Lecture Notes: Estimation of dynamic discrete choice models*

Jean-François Houde
Wharton School
University of Pennsylvania

March 16, 2015

*These lectures notes incorporate material from Victor Agguirregabiria’s graduate IO slides at the University of Toronto (http://individual.utoronto.ca/vaguirre/courses/eco2901/teaching_io_toronto.html).
Introduction

• The second part of the course deals with dynamic structural models in IO

• We start with an single-agent models of dynamic decisions:
  – Machine replacement and investment decisions: Rust (1987)
  – Renewal or exit decisions: Pakes (1986)
  – Inventory control: Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a)

This lecture will focus on econometrics methods, and next lecture will discuss mostly applications.

• Next, we will discuss questions related to the dynamic of industries:
  – “Macro” models of industry dynamics
  – Markov-perfect dynamic games
  – Empirical model of dynamic games
Machine replacement and investment decisions

• Consider a firm producing a good at $N$ plants (indexed by $i$) that operate independently. Each plant has a machine.

• Examples:
  – Rust (1987): Each plant is a Madison WI bus, and Harold Zucher is the plant operator.
  – Das (1992): Consider cement plants, where the machines are cement kiln.
  – Related applications: Export decisions (Das et al. (2007)), replacement of durable goods (Adda and Cooper (2000), Gowrisankaran and Rysman (2012)).
• Profit function at time $t$:

$$\pi_t = \sum_{i=1}^{N} y_{it} - r_{cit}$$

where $y_{it}$ is the plant’s variable profit, and $r_{cit}$ is the replacing cost of the machine.

• Replacement and depreciation:

  – Replace cost:

    $$r_{cit} = a_{it} \times RC(x_{it})$$

    where $\partial RC(x)/\partial x \geq 0$ and $a_{it} = 1$ if the machine is replaced. In the application, $RC(x_{it}) = \theta R_0 + \theta R_1 x_{it}$.

  – State variable: machine age $x_{it}$, choice-specific profit shock $\{\epsilon_{it}(0), \epsilon_{it}(1)\}$.

  – Variable profits are decreasing in the age $x_{it}$ of the aging, and increasing in profit shock $\epsilon_{it}(a_{it})$:

    $$y_{ij} = Y((1 - a_{it})x_{it}, \epsilon_{it}(a_{it}))$$

    where $\partial Y/\partial x < 0$.

  – This transforms the variable profit into a step function:

    $$\pi_{it} = \begin{cases} 
    Y(0, \epsilon_{it}(1)) - RC(x_{it}) & \text{If } a_{it} = 1 \\
    Y(x_{it}, \epsilon_{it}(0)) & \text{Otherwise.} 
    \end{cases}$$

  – Aging/depreciation process:

    Deterministic: $x_{it+1} = (1 - a_{it})x_{it} + 1$

    Stochastic: $x_{it+1} = (1 - a_{it})x_{it} + \xi_{t+1}$

In Rust (1987), $x_{it}$ is bus mileage. It follows a random walk process with a log-normal distribution.
Assumptions:

– Additive separable (AS) profit shock:

\[ Y((1 - a)x, \epsilon(a)) = \theta Y_0 + \theta Y_1 (1 - a)x + \epsilon(a) \]

– Conditional independence (CI): \( f(\epsilon_{t+1} | \epsilon_t, x_t) = f(\epsilon_{t+1}) \)

– Aging follows is a discrete random-walk process: \( x_{it} \in \{0, 1, ..., M\} \) and matrix \( F(x' | x, a) \) characterizes its controlled Markov transition process.

Dynamic optimization:

– Harold Zucher maximizes expected future profits:

\[
V(a_{it} | x_{it}, \epsilon_{it}) = E \left( \sum_{\tau=0}^{\infty} \beta^\tau \pi_{it+\tau} \middle| x_{it}, \epsilon_{it}, a_{it} \right)
\]

– Recursive formulation: Bellman equation

\[
V(a | x, \epsilon) = Y((1 - a) \cdot x) - RC(a \cdot x) + \epsilon(a) + \beta \sum_{x'} E_{\epsilon'}(V(x', \epsilon')) F(x' | x, a) = v(a, x) + \epsilon(a)
\]

where \( V(x, \epsilon) \equiv \max_{a \in \{0, 1\}} V(a | x, \epsilon) \).

– Optimal replacement decision:

\[
a^* = \begin{cases} 
1 & \text{If } v(1, x) - v(0, x) = \tilde{v}(x) > \epsilon(0) - \epsilon(1) = \tilde{\epsilon} \\
0 & \text{Otherwise.}
\end{cases}
\]

– If \( \{\epsilon(0), \epsilon(1)\} \) are distributed according to a type-1 EV distribution with unit variance:

\[
\Pr(a_{it} = 1 | x_{it}) = \exp(\tilde{v}(x_{it})/(1 + \exp(\tilde{v}(x_{it}))))
\]

\[
\tilde{V}(x_{it}) = E \left( \max_{a_{it}} v(a_{it}, x_{it}) + \epsilon_{it}(a_{it}) \right) = \ln \left( \sum_{a=0,1} \exp(v(a, x_{it})) \right)
\]

5
• Solution to the dynamic-programming (DP) problem:
  
  - Because of the additive-separability and conditional-independence assumptions we only need to numerically find a fixed-point to the "Emax" function:

    \[
    \bar{V}(x) = E_\epsilon \left( \max_a v(a, x) + \epsilon(a) \right)
    = E_\epsilon \left( \max_a \Pi(a, x) + \beta \sum_{x'} \bar{V}(x') F(x'|x, a) + \epsilon(a) \right)
    = \Gamma(x|\bar{V})
    \]

    where \(\Pi(a, x) = Y((1 - a) \cdot x) - RC(a \cdot x)\), and \(\Gamma(x|\bar{V})\) is a contraction mapping.

  - Matrix form representation using the EV distribution assumption:

    \[
    \bar{V} = \ln \left( \exp \left( \Pi(0) + \beta F(0) \bar{V} \right) + \exp \left( \Pi(1) + \beta F(1) \bar{V} \right) \right)
    = \Gamma(\bar{V})
    \]

    where \(F(0)\) and \(F(1)\) are two \(M \times M\) conditional transition probability matrix, and \(\bar{V}\) is the \(M \times 1\) vector of expected value function.

• Algorithm 1: Value-function iteration.

  it = 0 Guess initial value for \(\bar{V}(x)\). Example: \(\ln (\exp (\Pi(0)) + \exp (\Pi(1)))\).

  it = \(k\) Update value function:

    \[
    \bar{V}^k = \ln \left( \exp \left( \Pi(0) + \beta F(0) \bar{V}^{k-1} \right) + \exp \left( \Pi(1) + \beta F(1) \bar{V}^{k-1} \right) \right)
    \]

    Stop if \(||\bar{V}^k - \bar{V}^{k-1}|| < \eta\). Otherwise, update \(k + 1\).
• Algorithm 2: Policy-function iteration.

Define conditional choice-probability (CCP) function:

\[
P(x) = F_e \left( \Pi(1, x) + \beta \sum_{x'} \bar{V}(x') F(x'|x, 1) + \epsilon(1) \geq \Pi(0, x) + \beta \sum_{x'} \bar{V}(x') F(x'|x, 0) + \epsilon(0) \right) \\
= \exp(\bar{v}(x)/(1 + \exp(\bar{v}(x)))) = (1 + \exp(-\bar{v}(x)))^{-1}
\]

At the “optimal” CCP, we can write the Emax function as follows:

\[
\bar{V}^P(x) = (1 - P(x)) \left[ \Pi(0, x) + e(0, x) + \beta \sum_{x'} \bar{V}^P(x') F(x'|x, 0) \right] \\
+ P(x) \left[ \Pi(0, x) + e(1, x) + \beta \sum_{x'} \bar{V}^P(x') F(x'|x, 1) \right]
\]

where \( e(a, x) = E(\epsilon(a)|a, x) \) is the conditional expectation of the profit shock. If \( \epsilon(a) \) is EV distributed, we can write this expectation analytically:

\[
e(a, x) = \gamma - \ln P(a|x).
\]

This implicit define the value function in terms of the CCP vector:

\[
\bar{V}^P = (I - \beta F^P)^{-1} \left[ (1 - P) * (\Pi(0) + e(0)) + P * (\Pi(1) + e(1)) \right]
\]

where \( F^P = (1 - P) * F(0) + P * F(1) \).

Equations 1 and 2 define a fixed-point in \( P: P^* = \Psi(P^*) \). Moreover, \( \Psi(\cdot) \) is also a contraction mapping.
Algorithm steps:

it=0 Guess initial value for the CCP. Example: $P(x) = (1 + \exp(-(\Pi(1) - \Pi(0))))^{-1}$.

it=k Calculate expected value function:

\[
\tilde{V}^{k-1} = (I - \beta F^{k-1})^{-1} \left[ (1 - P^{k-1}) \ast (\Pi(0) + e^{k-1}(0)) + P^{k-1} \ast (\Pi(1) + e^{k-1}(1)) \right]
\]

Update CCP:

\[
P^{k} = \Psi(P^{k-1}) = (1 + \exp(-\tilde{v}^{k-1}))^{-1}
\]

where $\tilde{v}^{k-1} = (\Pi(1) + \beta F(1)\tilde{V}^{k-1}) - (\Pi(0) + \beta F(0)\tilde{V}^{k-1})$.

Stop if $||P^{k} - P^{k-1}|| < \eta$. Otherwise, update $k + 1$.

Note:

- Both algorithms are guaranteed to converge if $\beta \in (0, 1)$ (slower if $\beta \approx 1$).

- Policy-function iteration algorithms converges in fewer steps than value-function iteration.

- However, each step of the policy-function algorithm is \textbf{slower} due to the matrix inversion. $M$ is typically very large (in the millions).
• **Estimation:** Nested fixed-point maximum-likelihood (or NFXP)

  - **Data:** Panel of choices $a_{it}$ and observed states $x_{it}$
  - **Parameters:** Technology parameters $\theta = \{\theta_Y, \theta_R, \theta_{R_0}, \theta_{R_1}\}$, discount factor $\beta$, and distribution of mileage shocks $f_x(\xi_{it})$.
  - Likelihood contribution of bus $i$:

    $$L_i(A_i, X_i|\theta, \beta, f_x) = \prod_{t=1}^{T} \Pr(a_{it}|x_{it})f_x(x_{it} - a_{it-1}x_{it-1})$$

    where $\Pr(a_{it}|x_{it}) = \Psi(a_{it}|x_{it})$ is the solution to the DP problem.

  - If the panel is long-enough, we can efficiently estimate $f_x(\xi)$ from the data, and estimate the remaining parameters in a second stage (standard-errors should be adjusted however).

  - Maximum likelihood problem:

    $$\max_{\theta, \beta} \sum_i \sum_t a_{it} \ln P(x_{it}) + (1 - a_{it}) \ln (1 - P(x_{it}))$$

    s.t. $P(x_{it}) = \Psi(x_{it}) \ \forall x_{it}$

  - In practice, you want to solve the MLE problem by embedding the maximization problem into a standard optimization package (e.g. simplex or BFGS), and writing a fixed-point routine that solves the value-function or CCP contraction mapping. Using the matrix representation greatly helps if you use a matrix language like Matlab.
• Adding unobserved heterogeneity:
  – Assume that buses belong to one of $K$ discrete types
  – Example: Heterogenous replacement cost $\theta^k_{R_0}$
  – This increases the number of parameters by $K(K-1)$: $\{\theta^1_{R_0}, ..., \theta^K_{R_0}\}$
    $+ \{\omega_1, ..., \omega_{K-1}\}$ (probability weights).
  – This changes the MLE problem:

$$\max_{\theta, \beta, \omega} \sum_i \ln \left[ \sum_k \omega_k \prod_t P_k(x_{it})^{a_{it}} (1 - P_k(x_{it}))^{1-a_{it}} \right]$$

$$\text{s.t. } P_k(x_{it}) = \Psi_k(x_{it}) \quad \forall x_{it} \text{ and } k$$

This complicates the MLE significantly: Need to solve the DP problem separately for each type.
• **Identification:**

  – Standard normalization: $\sigma_\epsilon = 1$. This means that we cannot identify the “dollar” value of replacement costs. Only relative to variable profits.
  – When profits or output data are available, we can relax this normalization, and estimate $\sigma_\epsilon$ (e.g. investment and production data).
  – Discount factor?
    * Assume we know $F_\epsilon$.
    * $\beta$ is not identified, unless we impose parametric assumptions on $Y$ and $RC$.
    * The data corresponds to empirical hazard functions:
      
      $$h(x) = \Pr(\text{replacement}_t | \text{miles}_t = x)$$

    * This corresponds to the reduced form of the model:
      
      $$P(x) = F_\epsilon(\tilde{v}(x)) = F_\epsilon\left(-\beta \sum_{x'} V(x') (F(x'|1) - F(x'|0))\right)$$

    * If $\Pi(x)$ is linear in $x$, then any non-linearity in the observed hazard function identifies $\beta$.
    * If $\Pi(x)$ is allowed to take any non-linear function, we cannot distinguish between a non-linear myopic model ($\beta = 0$), and a forward-looking model ($\beta > 0$).
    * $\beta$ would be identified if the model included a state variable $z$ that only enters the Markov transition function (i.e. $F(x'|x, z, a)$), and not the static payoff function.
- Hazard function
Identification of \( \beta \) and search for the right specification:

<table>
<thead>
<tr>
<th>Cost Function</th>
<th>1, 2, 3</th>
<th>4</th>
<th>1, 2, 3, 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic ( c(x, \theta_1) = \theta_{11} x + \theta_{12} x^2 + \theta_{13} x^3 )</td>
<td>Model 1</td>
<td>Model 9</td>
<td>Model 17</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.063</td>
<td>-162.885</td>
<td>-296.515</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.177</td>
<td>-162.988</td>
<td>-296.411</td>
</tr>
<tr>
<td>quadratic ( c(x, \theta_1) = \theta_{11} x + \theta_{12} x^2 )</td>
<td>Model 2</td>
<td>Model 10</td>
<td>Model 18</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.326</td>
<td>-163.402</td>
<td>-297.939</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.534</td>
<td>-163.771</td>
<td>-299.328</td>
</tr>
<tr>
<td>linear ( c(x, \theta_1) = \theta_{11} x )</td>
<td>Model 3</td>
<td>Model 11</td>
<td>Model 19</td>
</tr>
<tr>
<td>\quad</td>
<td>-132.389</td>
<td>-163.584</td>
<td>-300.250</td>
</tr>
<tr>
<td>\quad</td>
<td>-134.747</td>
<td>-165.458</td>
<td>-306.641</td>
</tr>
<tr>
<td>square root ( c(x, \theta_1) = \theta_{11} \sqrt{x} )</td>
<td>Model 4</td>
<td>Model 12</td>
<td>Model 20</td>
</tr>
<tr>
<td>\quad</td>
<td>-132.104</td>
<td>-163.395</td>
<td>-299.314</td>
</tr>
<tr>
<td>\quad</td>
<td>-133.472</td>
<td>-164.143</td>
<td>-302.703</td>
</tr>
<tr>
<td>power ( c(x, \theta_1) = \theta_{11} x^{\theta_{12}} )</td>
<td>Model 5(^b)</td>
<td>Model 13(^b)</td>
<td>Model 21(^b)</td>
</tr>
<tr>
<td>\quad</td>
<td>N.C.</td>
<td>N.C.</td>
<td>N.C.</td>
</tr>
<tr>
<td>\quad</td>
<td>N.C.</td>
<td>N.C.</td>
<td>N.C.</td>
</tr>
<tr>
<td>hyperbolic ( c(x, \theta_1) = \theta_{11}/(91-x) )</td>
<td>Model 6</td>
<td>Model 14</td>
<td>Model 22</td>
</tr>
<tr>
<td>\quad</td>
<td>-133.408</td>
<td>-165.423</td>
<td>-305.605</td>
</tr>
<tr>
<td>\quad</td>
<td>-138.894</td>
<td>-174.023</td>
<td>-325.700</td>
</tr>
<tr>
<td>mixed ( c(x, \theta_1) = \theta_{11}/(91-x) + \theta_{12} \sqrt{x} )</td>
<td>Model 7</td>
<td>Model 15</td>
<td>Model 23</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.418</td>
<td>-163.375</td>
<td>-298.866</td>
</tr>
<tr>
<td>\quad</td>
<td>-131.612</td>
<td>-164.048</td>
<td>-301.064</td>
</tr>
<tr>
<td>nonparametric ( c(x, \theta_1) ) any function</td>
<td>Model 8</td>
<td>Model 16</td>
<td>Model 24</td>
</tr>
<tr>
<td>\quad</td>
<td>-110.832</td>
<td>-138.556</td>
<td>-261.641</td>
</tr>
<tr>
<td>\quad</td>
<td>-110.832</td>
<td>-138.556</td>
<td>-261.641</td>
</tr>
</tbody>
</table>

\(^a\) First entry in each box is (partial) log likelihood value \( \ell^2 \) in equation (5.2) at \( \beta = .9999 \). Second entry is partial log likelihood value at \( \beta = 0 \).

\(^b\) No convergence. Optimization algorithm attempted to drive \( \theta_{12} \to 0 \) and \( \theta_{11} \to +\infty \).
Main estimation results:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimates/Log-Likelihood</th>
<th>Data Sample</th>
<th>Heterogeneity Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta = .9999 )</td>
<td>RC</td>
<td>Groups 1, 2, 3 (3864 Observations)</td>
<td>Group 4 (4292 Observations)</td>
</tr>
<tr>
<td>( \theta_{11} )</td>
<td>11.7270 (2.602)</td>
<td>10.0750 (1.582)</td>
<td>9.7558 (1.227)</td>
</tr>
<tr>
<td>( \theta_{30} )</td>
<td>4.8259 (1.792)</td>
<td>2.2930 (0.639)</td>
<td>2.6275 (0.618)</td>
</tr>
<tr>
<td>( \theta_{31} )</td>
<td>.3010 (.0074)</td>
<td>.3919 (.0075)</td>
<td>.3489 (.0052)</td>
</tr>
<tr>
<td>( LL )</td>
<td>-.6884 (.0075)</td>
<td>-.5953 (.0075)</td>
<td>-.6394 (.0053)</td>
</tr>
<tr>
<td>( \beta = 0 )</td>
<td>RC</td>
<td>Groups 1, 2, 3 (1,0417)</td>
<td>Group 4 (7,6358 (0.7197))</td>
</tr>
<tr>
<td>( \theta_{11} )</td>
<td>8.2985 (1.6213)</td>
<td>7.1833 (1.3778)</td>
<td>7.0276 (1.0750)</td>
</tr>
<tr>
<td>( \theta_{30} )</td>
<td>.3010 (.0074)</td>
<td>.3919 (.0075)</td>
<td>.3488 (.0052)</td>
</tr>
<tr>
<td>( \theta_{31} )</td>
<td>.6884 (.0075)</td>
<td>.5953 (.0075)</td>
<td>.6394 (.0053)</td>
</tr>
<tr>
<td>( LL )</td>
<td>-.2710.746</td>
<td>-.3306.028</td>
<td>-.6061.641</td>
</tr>
</tbody>
</table>

Myopia test:

<table>
<thead>
<tr>
<th>LR Statistic ((df = 1))</th>
<th>Marginal Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.760</td>
<td>12.782</td>
</tr>
<tr>
<td>3.746</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

\( \beta = 0 \) vs. \( \beta = .9999 \)
This paper studies the value of patent protection: (i) what is the stochastic process determining the value of innovations?, (ii) how patent protection laws affect the decision to renew patents and the distribution of returns to innovation?

The model is an example of an **optimal stopping problem**. The model is setup with a finite horizon, but it does not have to be. Other examples: retirement, firm exit decisions, technology adoption, etc.

**Contributions:**

- Illustrate how we can infer the implicit option value of patents (or any other dynamic investment decision) from dynamic discrete choices (i.e. principle of revealed preference).
- This is done without actually observing profits or revenues from patents. Only the dynamic structure of renewal costs are needed.
- More technically, the paper is one of the first applications of simulation methods in econometrics (very influential).

**Data:**

- Three countries: France, Germany and UK
- Renewal date for all patents: \( n_{m,t}(a) \) = number of surviving patents at age \( a \) in country \( m \) from cohort \( t \).
- Regulatory environment by country/cohort:
  * \( f \): Number of automatic renewal years.
  * \( L \): Expiration date on patent
  * \( c = \{c_1, ..., c_T\} \): Deterministic renewal cost
<table>
<thead>
<tr>
<th>Country Characteristic</th>
<th>France</th>
<th>U.K.</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( f )</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>2. ( L )</td>
<td>20</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>4. First/last year in which renewals are observed</td>
<td>1970/81</td>
<td>1955/78</td>
<td>1955/74</td>
</tr>
<tr>
<td>5. Patents studied from cohort: all patents</td>
<td>Applied for</td>
<td>Applied for</td>
<td>Granted</td>
</tr>
<tr>
<td>6. Estimated average ratio of patents granted to</td>
<td>.93</td>
<td>.83</td>
<td>.35</td>
</tr>
<tr>
<td>patents applied for(^b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. ( \bar{NPAT} = N/J )</td>
<td>36,865</td>
<td>37,286</td>
<td>21,273</td>
</tr>
</tbody>
</table>
- Country differences in survival probabilities:

![Graph showing survival probabilities across different countries.](image)

**Figure 2.**—Average drop out proportions.

- Country differences in renewal fee schedules:

![Graph showing renewal fee schedules across different countries.](image)

**Figure 3.**—Average of renewal fee schedules.
• Model setup:
  
  – Consider the renewal problem for patent $i$
  
  – Stochastic sequence of returns from patent: $r_i = \{r_{i1}, ..., r_{iL}\}$
  
  – Evolution of returns depend on: (i) initial quality level, (ii) arrival of substitutes innovations that depreciate the value of the patent, (iii) arrival of complement innovations that increase its value.
  
  – Markov process for returns:

    \[
    r_{it+1} = \tau_{it+1} \max\{\delta r_{it}, \xi_{it+1}\}
    \]

    Where,

    \[
    \Pr(\tau_{it+1} = 0|r_{it}, t) = \exp(-\lambda r_{it})
    \]

    \[
    p(\xi_{it+1}|r_{it}, t) = \frac{1}{\phi t \sigma} \exp\left(-\frac{\gamma + \xi_{it+1}}{\phi t \sigma}\right)
    \]

    \[
    r_{i0} \sim LN(\mu_0, \sigma_0^2)
    \]

    or more compactly for $t > 0$,

    \[
    f(r_{it+1}|r_{it}, t) = \begin{cases} 
    \exp(-\lambda r_{it}) & \text{If } r_{it+1} = 0 \\
    \Pr(\xi_{it+1} < \delta r_{it}|r_{it}, t) & \text{If } r_{it+1} = \delta r_{it} \\
    \frac{1}{\phi t \sigma} \exp\left(-\frac{\gamma + \xi_{it+1}}{\phi t \sigma}\right) & \text{If } r_{it+1} > \delta r_{it}
    \end{cases}
    \]

  – Model structural parameters (per country):

    * $\delta$ measures the normal obsolescence rate
    
    * $\phi$ and $\sigma$ determines the arrival rate and magnitude of complementary innovations
    
    * $\lambda$ determines to arrival rate of substitute innovations
    
    * $\mu_0$ and $\sigma_0$ determines the initial quality pool of innovations
    
    Discount factor $\beta$ is fixed.
– Optimal stopping problem:

* In the last year, the value of renewing the patent depends only on $c_L$ and $r_{iL}$:

$$V(L, r_{iL}) = \max\{0, r_{iL} - c_L\}$$

and therefore the patent is renewed if $r_{iL} > r_{L}^* = c_L$.

* At year $L - 1$, the value is defined recursively:

$$V(L, r_{iL-1}) = \max\{0, r_{iL-1} - c_{L-1} + \beta \int_{r_{L-1}^*}^{\infty} V(L, r_{iL}) f(r_{iL} | r_{iL-1}, L - 1) dr_{iL}\}$$

This value function is strictly increasing in $r_{iL-1}$ (see proposition 1). Therefore, there exists a unique threshold such that the patent is renewed if

$$r_{iL-1} > r_{L-1}^* = c_{L-1} - \beta \int_{r_{L-1}^*}^{\infty} V(L, r_{iL}) f(r_{iL} | r_{L-1}^*, L - 1) dr_{iL}$$

* Similarly, for any year $t > 0$ the value function is defined recursively as follows:

$$V(L, r_{it}) = \max\{0, r_{it} - c_t + \beta \int_{r_{it+1}^*}^{\infty} V(t + 1, r_{it+1}) f(r_{it+1} | r_{it}, t) dr_{it+1}\}$$

which lead to a series of optimal stopping rules:

$$r_{it} > r_t^* = c_t - \beta \int_{r_t^*}^{\infty} V(t + 1, r_{it+1}) f(r_{it+1} | r_t^*, t)$$

* Given the function form assumptions on $f(r' | r_t, t)$, the thresholds can be solved analytically by backward induction.

* When the terminal period is stochastic the value function becomes stationary. For instance, optimal stopping problems arise when studying retirement or exit decisions:

$$V(s_t) = \max \left\{ 0, \pi(s_t) + \beta \int (1 - \delta(s_t)) V(s_{t+1}) f(s_{t+1} | s_t) ds_{t+1} \right\}$$
- **Estimation:**

  - Likelihood of the observed renewal sequence $N_m$ conditional on the regulation environment $Z_m = \{L_m, f_m, c_m\}$ in country $m$:

    $$L(N_m|Z_m, \theta) = \max_\theta \sum_{t=1}^L n_m(t) \ln \Pr(t^* = t|Z_m, \theta)$$

    Where,

    $$\Pr(t^* = t|\theta, Z_m) = \int_{-\infty}^{\infty} \int_{r_1^*}^{\infty} \int_{r_2^*}^{\infty} \cdots \int_{r_t^*}^{\infty} dF(r_{i1}, \ldots, r_{it-1}, r_{it})dF_0(r_{i0})$$

  - Monte Carlo integration approximation:

    0. Sample $r_{i0}^s \sim LN(\mu_0, \sigma_0^2)$

    1. Period 1:

      (a) Sample $\tau_1^s$ from Bernoulli with probability $\exp(-\lambda r_{i0}^s)$

      (b) If $\tau_1^s = 1$, sample $\xi_1^s$ from exponential distribution. Otherwise, do not renew patent: $a_1^s = 0$.

      (c) Calculate $r_1^1$

      (d) Evaluate decision: $a_1^s = 1$ if $r_1^s > r_1^*$.

    ...

    t. Repeat sampling for period $t$ if patent was renewed at $t - 1$

    After collecting the simulated sequences of actions, we can evaluate the simulated choice-probability at period $t$:

    $$\tilde{P}_S(t|\theta, Z_m) = \frac{1}{S} \sum_s 1(a_1^s = 1, a_2^s = 1, \ldots, a_{t-1}^s = 1, a_t^s = 0)$$

    - Numerical problem: $\tilde{P}_S(t|\theta, Z_m)$ is not a smooth function of the parameters $\theta +$ equal to zero for some $t$ unless $S \to \infty$.

    - Smooth alternative approximation:

      $$\hat{P}_S(t, \theta, Z_m) = \frac{\exp \left( \tilde{P}_S(t|\theta, Z_m)/\eta \right)}{1 + \sum_{t' \neq t} \exp(\tilde{P}_S(t'|\theta, Z_m)/\eta)}$$
– **Note:** All the structural parameters are identified in this model. The implicit normalization is that coefficient on renewal cost $c_t$: all the parameters are expressed in dollar.

- **Results:**

  - Main differences across countries: (i) patent regulation rules, (ii) initial distribution of patent returns.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>PARAMETER ESTIMATES*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>France</td>
</tr>
<tr>
<td>A. Parameter</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>5689 (8.24)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>9162 (13.67)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>.5084 ($5.66 \times 10^{-4}$)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>.8475 ($2.62 \times 10^{-4}$)</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>1.579 ($2.92 \times 10^{-3}$)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>4.705 ($2.75 \times 10^{-3}$)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>.0990 ($6.36 \times 10^{-4}$)</td>
</tr>
</tbody>
</table>

B. Dimension

<table>
<thead>
<tr>
<th>Country</th>
<th>France</th>
<th>U.K.*</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1. NPAT</td>
<td>1,069,095</td>
<td>983,471</td>
<td>446,741</td>
</tr>
<tr>
<td>B.2. NSIM</td>
<td>20,000</td>
<td>20,000</td>
<td>20,000</td>
</tr>
<tr>
<td>B.3. Age: $f/L$</td>
<td>2/20</td>
<td>5/16</td>
<td>3/18</td>
</tr>
<tr>
<td>B.4. NCHRT</td>
<td>29</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>B.5. NCHRTAGE</td>
<td>238</td>
<td>272</td>
<td>237</td>
</tr>
</tbody>
</table>

C. Summary Statistic

<table>
<thead>
<tr>
<th>Country</th>
<th>France</th>
<th>U.K.*</th>
<th>Germany</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1. MSE[$\hat{\pi}$]</td>
<td>$5.42 \times 10^{-4}$</td>
<td>$6.91 \times 10^{-4}$</td>
<td>$1.48 \times 10^{-4}$</td>
</tr>
<tr>
<td>C.2. PDW[$\hat{\pi}$]</td>
<td>1.65</td>
<td>2.24</td>
<td>1.85</td>
</tr>
<tr>
<td>C.3. $V[\hat{\pi};$ data]</td>
<td>$3.90 \times 10^{-2}$</td>
<td>$1.07 \times 10^{-2}$</td>
<td>$2.65 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

*Patents are assigned to cohorts by year of application. Numbers in parenthesis beside parameter estimates are their estimated standard errors.

Germany has a more selective screening system for granting new patents: higher mean and smaller variance of initial returns $r_{i0}$.
Learning about complementary innovations: $\phi \approx 0.5$. Imply very fast learning/growth in returns.

This has important policy implications: Regulator wants to keep initial renewing cost low, and increase them fast to extract rents from high value patents (low distortions after learning is over).

![Table III: The Evolution of Implicit Revenues in the Early Ages](image)

This table shows the evolution of implicit revenues for France and Germany in the early ages. The table includes various characteristics and their corresponding values for each country.
- The distribution of realized patent value is highly skewed:

| Per cent | France (p| $) | lc per cent | U.K. (pl $) | lc per cent | Germany (pl $) | lc per cent |
|----------|--------|-------------|-------------|-------------|---------------|-------------|
| .25      | 75.23  | .544        | 355.55      | .554        | 1,999.60      | 2.249       |
| .50      | 533.96 | 1.833       | 1,516.84    | 3.247       | 6,252.93      | 7.341       |
| .75      | 3,731.35 | 8.087   | 7,947.55    | 16.369      | 19,576.26     | 25.288      |
| .85      | 10,292.06 | 19.575  | 15,357.09   | 31.721      | 32,428.14     | 41.001      |
| .90      | 17,423.11 | 31.261  | 22,206.21   | 44.257      | 44,241.87     | 52.672      |
| .95      | 31,609.59 | 52.461  | 34,740.07   | 62.960      | 65,753.61     | 69.223      |
| .97      | 42,905.78 | 65.514  | 43,889.95   | 73.640      | 78,299.01     | 78.348      |
| .98      | 51,215.84 | 73.729  | 51,277.22   | 80.072      | 94,842.63     | 83.800      |
| .99      | 66,515.40 | 84.011  | 65,075.08   | 87.858      | 118,354.78    | 90.330      |
| maximum  | 259,829.27 | —       | 374,028.70  | —           | 419,217.55    | —           |
| mean     | 5,631.03  | —         | 7,357.05    | —           | 16,169.48     | —           |
| NPAT     | 36,865   | 37,826     | 21,273      |             |               |             |

* The realized value for patent $i$ is $\sum_{t=1}^{T_i} \beta^{(t-1)}(r_{it} - c_i)$, where $T_i$ is the last age at which patent $i$ was kept in force. See also the note to Table III.

- Implied rate of returns on R&R: France = 15.56%, UK = 111.03%, Germany = 13.83%.
Sequential estimators of DDC models

- **Key references:**
  - Hotz and Miller (1993)
  - Hotz, Miller, Sanders, and Smith (1994)
  - Aguirregabiria and Mira (2002)
  - Identification issues: Magnac and Thesmar (2002), Kasahara and Shimotsu (2009)

- Consider the following dynamic discrete choice model with additively separable (AS) and conditional independent (CI) errors.
  - A discrete actions.
  - Payoff function: $u(x|a)$
  - State space: $(x, \epsilon)$.
  - Where $x$ is a discrete state vector, and $\epsilon$ is a $A$ continuous vector.
  - Distribution functions:
    * $\Pr(x_{t+1} = x'|x_t, a) = f(x'|x, a)$
    * $g(\epsilon)$ is a type-1 EV density with unit variance.

- **Bellman operator:**

  $$V(x) = \int \max_{a \in A} \left\{ u(x|a) + \epsilon(a) + \beta \sum_{x'} V(x') f(x'|x, a) \right\} g(\epsilon) d\epsilon$$

  $$= \int \max_{a \in A} \left\{ v(x|a) + \epsilon(a) \right\} g(\epsilon) d\epsilon$$

  $$= \ln \left( \sum_{a} \exp(v(x|a)) \right)$$

  $$= \Gamma(V(x))$$
CCP operator:

* Express \( V(x) \) as a function of \( P(a|x) \).

\[
V(x) = \sum_a P(a|x) \ast \left\{ u(x|a) + E(\epsilon(a)|x,a) + \beta \sum_{x'} V(x') f(x'|x,a) \right\}
\]

Where,

\[
E(\epsilon(a)|x,a) = \frac{1}{P(a|x)} \int 1(v(x|a) + \epsilon(a) > v(x|a') + \epsilon(a'), a' \neq a) g(\epsilon) d\epsilon
\]

\[
e(a, P(a|x)) = \gamma - \ln P(a|x)
\]

* In Matrix form:

\[
V = \sum_a P(a) \ast \left[ u(a) + e(a, P) + \beta F(a)V \right]
\]

\[
\left[ I - \beta \sum_a P(a) \ast F(a) \right] V = \sum_a P(a) \ast \left[ u(a) + e(a, P) \right]
\]

\[
V(P) = \left[ I - \beta \sum_a P(a) \ast F(a) \right]^{-1} \left[ \sum_a P(a) \ast (u(a) + e(a, P)) \right]
\]

where \( F(a) \) is \( |X| \times |X| \) and \( V \) is \( |X| \times 1 \).

* The CCP contraction mapping is:

\[
P(a|x) = \text{Pr} \left( v(x|a, P) + \epsilon(a) > v(x|a', P) + \epsilon(a'), a' \neq a \right)
\]

\[
= \frac{\exp \left( \tilde{v}(x|a, P) \right)}{1 + \sum_{a' > 1} \exp \left( \tilde{v}(x|a, P) \right)}
\]

\[
= \Psi(a|x, P)
\]

where \( \tilde{v}(x|a, P) = v(x|a, P) - v(x|1, P) \).

**Special case 1:** If \( u(x|a, \theta) = x\theta \), the value function is also linear in \( \theta \).

\[
V(P) = Z(P)\theta + \lambda(P)
\]

Where

\[
Z(P) = \left[ I - \beta \sum_a P(a) \ast F(a) \right]^{-1} \left[ \sum_a P(a) \ast X \right]
\]

\[
\lambda(P) = \left[ I - \beta \sum_a P(a) \ast F(a) \right]^{-1} \left[ \sum_a P(a) \ast e(a, P) \right]
\]
• **Special case 2:** Absorbing state, such that \( v(x|0) = 0 \) (e.g. Exit or retirement). This change the value function:

\[
V(x, \varepsilon) = \max \left\{ u(x) + \varepsilon(1) + \beta \sum_{x'} \mathbb{E}_{\varepsilon'}[V(x', \varepsilon')] F(x'|x), \varepsilon(0) \right\}
\]

As before, the expected continuation value is:

\[
\tilde{V}(x) = \log \left( \exp(0) + \exp \left( u(x) + \beta \sum_{x'} \tilde{V}(x') F(x'|x) \right) \right) + \gamma
\]

\[
= \log \left( 1 + \exp(v(x)) \right) + \gamma
\]

where \( v(x) = u(x) + \beta \sum_{x'} \tilde{V}(x') \).

The choice probability is given by:

\[
\Pr(a = 1|x) = P(x) = \frac{\exp(v(x))}{1 + \exp(v(x))}
\]

Note that the log of the “odds-ratio” is equal to the choice-specific value function:

\[
\log \left( \frac{P(x)}{1 - P(x)} \right) = v(x)
\]

Therefore, the expected continuation value can be expressed as a function of \( P(x) \):

\[
\tilde{V}^p(x) = \log \left( 1 + \exp(v(x)) \right) + \gamma = \log \left( 1 + \frac{P(s)}{1 - P(x)} \right) + \gamma
\]

\[
= -\log \left( 1 - P(x) \right) + \gamma
\]

Therefore, with an absorbing state, we don’t need to invert \([I - \beta \sum_a P(a) F(a)]\) to apply the CCP mapping.
• **Two-step Estimator:**
  
  – The objective is to estimate the structural parameters $\theta$ without repeatedly solving the DP problem
  
  – **Initial step:** Reduced form of the model
    
    * Markov transition process: $\hat{f}(x'|x,a)$
    * Policy function: $\hat{P}(a|x)$
    * **Constraint:** Need to estimate both functions at EVERY state point $x$.

  – How? Ideally $\hat{P}(a|x)$ is estimated non-parametrically to avoid imposing a particular functional form on the policy function (i.e. no theory involved at this stage). This would correspond to a frequency estimator:

    $$
    \hat{P}(a|x) = \frac{1}{n(x)} \sum_{i \in n(x)} 1(a_i = a)
    $$

  – In practice, for finite samples, we need to impose smooth the policy function and interpolate between states are not visited (or infrequently). Kernels or local-polynomial techniques can be used.

  – **Second-step:** Structural parameters conditional on $(\hat{P}, \hat{f})$
− **Example:** Assume that \( u(x|a, \theta) = x(a)\theta \) and fix \( \beta \).

1- **Data Preparation:** Use \((\hat{P}, \hat{F})\) to calculate:

\[
Z(\hat{P}, \hat{F}) = \left[ I - \beta \sum_a \hat{P}(a) \ast \hat{F}(a) \right]^{-1} \left[ \sum_a \hat{P}(a) \ast X(a) \right]
\]

\[
\lambda(\hat{P}, \hat{F}) = \left[ I - \beta \sum_a \hat{P}(a) \ast F(a) \right]^{-1} \left[ \sum_a \hat{P}(a) \ast e(a, \hat{P}) \right]
\]

2- **GMM:** Let \( W_{it} \) denote a vector of predetermined instruments (e.g. state-variables and their interactions). We can construct a set of moment conditions:

\[
E \left( W_{it} \left[ a_{it} - \Psi(a_{it}|x_{it}, \hat{P}, \hat{F}) \right] \right) = 0
\]

Where,

\[
\Psi(a_{it}|x_{it}, \hat{P}, \hat{F}) = \frac{\exp \left( v(x_{it}|a_{it}, \hat{P}, \hat{F}) \right)}{\sum_{a'} \exp(v(x_{it}|a', \hat{P}, \hat{F}))}
\]

\[
v(x|a, \hat{P}, \hat{F}) = x(a)\theta + \beta \sum_{x'} \underbrace{V(x|\hat{P}, \hat{F})}_{=Z(x,\hat{P},\hat{F})} f(x'|x, a).
\]

\[
= \left( x(a) + \beta \hat{Z}(x|\hat{P}, \hat{F}) \right) \theta + \beta \hat{\lambda}(x|\hat{P}, \hat{F})
\]

Therefore, the second-stage of problem is equivalent to a linear IV problem.
- **Pseudo-likelihood estimators (PML):** Use the CCP mapping $\Psi(P)$ to construct a pseudo-likelihood estimator (Aguirregabiria and Mira (2002))

  - **Data:** Panel of $n$ individuals of $T$ periods:
    \[
    (A, X) = \{a_{it}, x_{it}\}_{i=1,...,n,t=1,...,T}
    \]

  - **2-Step estimator:**
    1. Obtain a flexible estimator of CCPs $\hat{P}^1(a|x)$
    2. Feasible PML estimator:
    \[
    Q^{2S}(A, X) = \max_{\theta} \sum_i \sum_t \Psi(a_{it}|x_{it}, \hat{P}^1, \hat{F}, \theta)
    \]
    If $V(P)$ is linear, the second step is a linear probit/logit model.

  - **NPL estimator:** The NPL repeat the PML and policy function iteration steps sequentially (i.e. swapping the fixed-point algorithm).
    0 Obtain a flexible estimator of CCPs $\hat{P}^1(a|x)$
    1 Feasible PML step:
    \[
    Q^{k+1}(A, X) = \max_{\theta} \sum_i \sum_t \Psi(a_{it}|x_{it}, \hat{P}^k, \hat{F}, \theta)
    \]
    2 Policy function iteration step:
    \[
    \hat{P}^{k+1}(a|x) = \Psi(a|x, \hat{P}^k, \hat{F}, \hat{\theta}^{k+1})
    \]
    3 Stop if $||\hat{P}^{k+1} - \hat{P}^k|| < \delta$, else repeat step 1 and 2.

- In the single agent case: The NPL is guaranteed to converge to the MLE estimator (i.e. NFXP).
- In practice, Aguirregabiria and Mira (2002) showed that 2 or 3 steps is sufficient to eliminate the small sample bias of the 2-step estimator, and is computationally easier to implement than the NFXP.
Simulation-based CCP estimator: Hotz, Miller, Sanders, and Smith (1994)

- **Starting point:** The H&M GMM estimator suffers from a curse of dimensionality in $|X|$, since we must invert a $|X| \times |X|$ matrix to evaluate the continuation value (not true for optimal-stopping models). This is less severe for NFXP estimators, since we can use the value-function mapping to solve the policy functions.

- **Solution:**
  
  * First insight: We only need to know the relative choice-specific value function $\tilde{v}(a|x) = v(a|x) - v(1|x)$ to predict behavior.

  $$a_{it} = \begin{cases} 
  1 & \text{If } \tilde{v}(a|x) + \tilde{\epsilon}(a) < 0 \text{ for all } a \neq 1 \\
  a & \text{If max}\{0, \tilde{v}(a'|x) + \tilde{\epsilon}(a')\} < \tilde{v}(a|x) + \tilde{\epsilon}(a) \text{ for all } a' \neq a
  \end{cases}$$

  * Second insight: There exists a one-to-one mapping between $\tilde{v}(a|x)$ and $P(a|x)$.

  Logit example:

  $$P(a|x) = \frac{\exp(v(a|x))}{\sum_a \exp(v(a'|x))} = \frac{\exp(\tilde{v}(a|x))}{1 + \sum_{a'>1} \exp(\tilde{v}(a'|x))}$$

  $$\Leftrightarrow \tilde{v}(a|x, P) = \ln P(a|x) - \ln P(1|x)$$

  * Third insight: We can approximate the model predicted value function at any state $x$ by simulating actions according to a policy function $P(a|x)$.

  $$\hat{V}^S(x|P) = \frac{1}{S} \sum_s \sum_{\tau=0}^T \beta^\tau \{ u(x_{t+\tau}^s, a_{t+\tau}^s) + e(a_{t+\tau}^s|P(a_{t+\tau}^s|x_{t+\tau}^s)) \}$$

  where $(x^s, a^s)$ is a simulated sequence of choices and states sampled from $P(a|x)$ and $f(x'|x, a)$, and $e(a|P(a|x)) = E(\epsilon(a)|a_i = a, x, P)$. Importantly, $\lim_{S \to \infty} \hat{V}^S(x|P) = V(s|P)$. 

30
– Estimation procedure:

* **Step 1:** Estimate \( \hat{P}(a|x) \) and \( \hat{f}(x'|x,a) \), and compute the “dependent variable”:

\[
\tilde{v}_n(a_{it}|x_{it}, \hat{P}) = \ln \hat{P}(a_{it}|x_{it}) - \ln \hat{P}(1|x_{it})
\]

* **Step 2a:** Simulation of value functions of each observed state and choice \((x_{it}, a_{it})\). Each simulated sequence calculate the value of “future” choices:

1. Calculate static value of \((x_{it}, a_{it})\):
   \[u(x_{it}, a_{it}|\theta) + e(a_{it}|\hat{P}, x_{it})\]
2. Sample new state for period \(t + 1\): \(x_{it+1} \sim \hat{f}(x'|x_{it}, a_{it})\)
3. Sample new choice for period \(t + 1\): \(a_{it+1} \sim \hat{P}(a|x_{it})\)

Repeat steps 1-3 for \(T\) periods. This gives us the net present value of one simulated sequence:

\[
v^s(a_{it}|x_{it}, \hat{P}, \theta) = u(x_{it}, a_{it}|\theta) + e(a_{it}|\hat{P}, x_{it+\tau})
+ \sum_{\tau=1}^{T} \beta^\tau \left[u(x^s_{it+\tau}, a^s_{it+\tau}|\theta) + e(a^s_{it+\tau}|\hat{P}, x^s_{it+\tau})\right]
\]

Repeat this process \(S\) times. This gives us the simulated value of choosing \(a_{it}\) in state \(x_{it}\):

\[
v^S(a_{it}|x_{it}, \hat{P}, \theta) = \frac{1}{S} \sum_s v^s(a_{it}|x_{it}, \hat{P})
\]

Let \(\tilde{v}^S(a_{it}|x_{it}, \hat{P}, \theta) = v^S(a_{it}|x_{it}, \hat{P}, \theta) - v^S(1|x_{it}, \hat{P}, \theta)\).

**Note:** If \(u(x, a|\theta)\) is linear in \(\theta\), we need to do this simulation process only once.

* **Step 2b:** Moment conditions

\[
E\left(W_{it} \left[\tilde{v}_n(a_{it}|x_{it}, \hat{P}) - \tilde{v}^S(a_{it}|x_{it}, \hat{P}, \theta)\right]\right) = 0
\]

where \(W_{it}\) is a vector of instruments.
* Importantly, setting up the moment conditions this way implies that the estimate will be consistent even with a finite number of simulated number of draws $S$.
* Why? The simulation error, $\tilde{v}(a_{it}|x_{it}, \hat{P}, \theta) - \tilde{v}^S(a_{it}|x_{it}, \hat{P}, \theta)$, is additive, and therefore vanishes as $n \to \infty$ (instead of $S \to \infty$).
* However, the small sample bias in $\hat{P}$ enters non-linearly in the moment conditions, and can induce severe biases:

\[
\ln \left( \hat{P}(a_{it}|x_{it}) + u_{it}(a) \right) - \ln \left( \hat{P}(1|x_{it}) + u_{it}(1) \right) \neq \ln \hat{P}(a_{it}|x_{it}) - \ln \hat{P}(1|x_{it}) + u_{it}
\]

For instance, if $\hat{P}(a_{it}|x_{it}) = 0$, the objective function is not defined.
* HMSS presents Monte-Carlo experiment to illustrate the small-sample bias. It can be quite large.
Demand for storable goods

• **Key references:**
  – Pesendorfer (2002)
  – Hendel and Nevo (2006a) and Hendel and Nevo (2006b)

• If consumers can stockpile a storable good, they can benefit from “sales”: purchase more than consumption when prices are low.

• This creates a difference between *purchasing* and *consumption* elasticity
  – In the short-run, store-level demand elasticity reflect stockpiling behavior
  – In the long-run, store-level demand elasticity reflects consumption decisions

• Ignoring stockpiling and forward-looking behavior can lead to biases: own and cross price elasticities.

• If consumers differ in the storage costs firms have an incentive to price discriminate: PD theory of sales.

• Ketchup pricing example (Pesendorfer 2002):

![Graph showing prices for selected products in supermarket 1](image)

*Fig. 1.—Prices for selected products in supermarket 1*
• **Hendel and Nevo (2006a):**
  - Estimate an inventory control model *with* product differentiation
  - Two challenges: (i) inventories are unobserved (key state variable), and (ii) differentiation creates a dimensionality problem
  - Propose a simplifying assumption that: (i) reduce the computation burden of the model (reduce the dimension), and (ii) simplify the estimation (three-step algorithm).

• **Data:**
  - Scanner data: Household panel with purchasing decisions from 1991-1993
  - Firms: 9 supermarkets (ignore supermarket choice)
  - Store-level data: choice-set and product attributes (detergent), promotions, and prices

```
<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMMARY STATISTICS OF HOUSEHOLD-LEVEL DATA*</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>Income (000’s)</td>
</tr>
<tr>
<td>Size of household</td>
</tr>
<tr>
<td>Live in suburb</td>
</tr>
<tr>
<td>Purchase of laundry detergents</td>
</tr>
<tr>
<td>Price ($)</td>
</tr>
<tr>
<td>Size (oz.)</td>
</tr>
<tr>
<td>Quantity</td>
</tr>
<tr>
<td>Duration (days)</td>
</tr>
<tr>
<td>Number of brands bought over the 2 years</td>
</tr>
<tr>
<td>Brand HHI</td>
</tr>
<tr>
<td>Store visits</td>
</tr>
<tr>
<td>Number of stores visited over the 2 years</td>
</tr>
<tr>
<td>Store HHI</td>
</tr>
</tbody>
</table>
```

*For Demographics, Store visits, Number of brands, and Brand HHI, an observation is a household. For all other statistics, an observation is a purchase instance. Brand HHI is the sum of the square of the volume share of the brands bought by each household. Similarly, Store HHI is the sum of the square of the expenditure share spent in each store by each household.
- Market shares and sales:

<table>
<thead>
<tr>
<th>Brand</th>
<th>Firm</th>
<th>Share</th>
<th>Cumulative</th>
<th>% on Sale</th>
<th>Brand</th>
<th>Firm</th>
<th>Share</th>
<th>Cumulative</th>
<th>% on Sale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tide</td>
<td>P &amp; G</td>
<td>21.4</td>
<td>21</td>
<td>32.5</td>
<td>Tide</td>
<td>P &amp; G</td>
<td>40</td>
<td>40</td>
<td>25.1</td>
</tr>
<tr>
<td>All</td>
<td>Unilever</td>
<td>15</td>
<td>36</td>
<td>47.4</td>
<td>Cheer</td>
<td>P &amp; G</td>
<td>14.7</td>
<td>55</td>
<td>9.2</td>
</tr>
<tr>
<td>Wisk</td>
<td>Unilever</td>
<td>11.5</td>
<td>48</td>
<td>50.2</td>
<td>A &amp; H</td>
<td>C &amp; D</td>
<td>10.5</td>
<td>65</td>
<td>28</td>
</tr>
<tr>
<td>Solo</td>
<td>P &amp; G</td>
<td>10.1</td>
<td>58</td>
<td>7.2</td>
<td>Dutch</td>
<td>Dial</td>
<td>5.3</td>
<td>70</td>
<td>37.6</td>
</tr>
<tr>
<td>Purex</td>
<td>Dial</td>
<td>9</td>
<td>67</td>
<td>63.1</td>
<td>Wisk</td>
<td>Unilever</td>
<td>3.7</td>
<td>74</td>
<td>41.2</td>
</tr>
<tr>
<td>Cheer</td>
<td>P &amp; G</td>
<td>4.6</td>
<td>72</td>
<td>23.6</td>
<td>Oxydol</td>
<td>P &amp; G</td>
<td>3.6</td>
<td>78</td>
<td>59.3</td>
</tr>
<tr>
<td>A &amp; H</td>
<td>C &amp; D</td>
<td>4.5</td>
<td>76</td>
<td>21.5</td>
<td>Surf</td>
<td>Unilever</td>
<td>3.2</td>
<td>81</td>
<td>11.6</td>
</tr>
<tr>
<td>Ajax</td>
<td>Colgate</td>
<td>4.4</td>
<td>80</td>
<td>59.4</td>
<td>All</td>
<td>Unilever</td>
<td>2.3</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Dow Chemical</td>
<td>4.1</td>
<td>85</td>
<td>33.1</td>
<td>Dreft</td>
<td>P &amp; G</td>
<td>2.2</td>
<td>86</td>
<td>15.2</td>
</tr>
<tr>
<td>Surf</td>
<td>Unilever</td>
<td>4</td>
<td>89</td>
<td>42.5</td>
<td>Gain</td>
<td>P &amp; G</td>
<td>1.9</td>
<td>87</td>
<td>16.7</td>
</tr>
<tr>
<td>Era</td>
<td>P &amp; G</td>
<td>3.7</td>
<td>92</td>
<td>40.5</td>
<td>Bold</td>
<td>P &amp; G</td>
<td>1.6</td>
<td>89</td>
<td>1.1</td>
</tr>
<tr>
<td>Generic</td>
<td>0.9</td>
<td>93</td>
<td>0.6</td>
<td></td>
<td>Generic</td>
<td>0.7</td>
<td>90</td>
<td>16.6</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>—</td>
<td>0.2</td>
<td>93</td>
<td>0.9</td>
<td>Other</td>
<td>—</td>
<td>0.6</td>
<td>90</td>
<td>19.9</td>
</tr>
</tbody>
</table>

*Columns labeled Share are shares of volume (of liquid or powder) sold in our sample, columns labeled Cumulative are the cumulative shares, and columns labeled % on Sale are the percent of the volume, for that brand, sold on sale. A sale is defined as any price at least 5 percent below the modal price, for each UPC in each store. A & H = Arm & Hammer; P & G = Procter and Gamble; C & D = Church and Dwight.*
• Model setup:
  - $J$ differentiated brands
  - Weekly decisions: consumption $c_t$, purchases quantity $x_t \in \{0, 1, 2, 3, 4\}$, and brand $j_t$
  - Utility function:
    \[
    U(c_t, v_t, i_{t+1}) = \gamma \ln(c_t + v_t) - \left[ \delta_1 i_{t+1} + \delta_2 i_{t+1}^2 \right] + m_t
    \]
    where $m_t = \sum_{j,x} d_{jxt}(\beta a_{jxt} - \alpha p_{jxt} + \xi_{jxt} + \epsilon_{jxt})$ is the indirect utility from purchasing the outside good plus the value of the chosen brand.
  - Inventory transition:
    \[
    i_{t+1} = i_t - c_t + x_t
    \]
  - **Note:** The function $m_t$ implies that taste for brand only enters the purchasing stage, no the consumption stage. That is, the utility of consumption does not depend on the mix of brands in storage.
  - State variables: $s_t = \{i_t, v_t, p_t, a_t, \xi_t, \epsilon_t\}$.

• Two challenges:
  1. Unobserved state variable: $i_t$ is unobserved to the econometrician.
  2. Curse of dimensionality: $\{p_t, a_t, \xi_t, \epsilon_t\}$ is $4 \times 5 \times J$ dimension matrix.

**Note:** In Rust (1987), we also had unobserved state variables (i.e. logit errors). However, we assumed that it was conditionally independent of the observed state and separately additive, which allowed us to work with the Emax function. Not the case here: $i_t$ is serially correlated and non-separable.
• Model choice-probability:

\[
\Pr(d_{jxt} = 1|s_t) = \frac{\exp(\beta a_{jxt} - \alpha p_{jxt} + \xi_{jxt} + M(x, s_t))}{\sum_{x',j'} \exp(\beta a_{j'x't} - \alpha p_{j'x't} + \xi_{j'x't} + M(x', s_t))}
\]

Where,

\[
M(x, s_t) = \max_c \gamma \ln(c + v_t) - [\delta_1 i_{t+1} + \delta_2 i_{t+1}^2] + \beta E[V(s_{t+1})|x, c, s_t]
\]

s.t. \( i_{t+1} = i_t - c + x \)

• Note that conditional on purchasing size \( x \), the brand choice is a static problem:

\[
\Pr(d_{jt} = 1|s_t, x_t = x) = \frac{\exp(\beta a_{jxt} - \alpha p_{jxt} + \xi_{jxt} + M(x, s_t))}{\sum_j \exp(\beta a_{j'xt} - \alpha p_{j'xt} + \xi_{j'xt} + M(x, s_t))}
\]

This probability can be estimated by MLE: Multinomial logit model.

• This is the first-step of the estimation procedure.

• Endogenous prices and advertising? Brand/size fixed-effects, such that \( \xi_{jxt} = \xi_{jx} \).

<table>
<thead>
<tr>
<th>TABLE IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIRST STEP: BRAND CHOICE CONDITIONAL ON SIZE(^a)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Price</td>
</tr>
<tr>
<td>(0.022)</td>
</tr>
<tr>
<td>*Suburban dummy</td>
</tr>
<tr>
<td>(0.055)</td>
</tr>
<tr>
<td>*Nonwhite dummy</td>
</tr>
<tr>
<td>(0.075)</td>
</tr>
<tr>
<td>Large family</td>
</tr>
<tr>
<td>(0.080)</td>
</tr>
<tr>
<td>Feature</td>
</tr>
<tr>
<td>(0.095)</td>
</tr>
<tr>
<td>Display</td>
</tr>
<tr>
<td>(0.069)</td>
</tr>
</tbody>
</table>

\(^a\)Estimates of a conditional logit model. An observation is a purchase instance by a household. Options include only products of the same size as the product actually purchased. Asymptotic standard errors are shown in parentheses.
• How to deal with the dimensionality problem?

• Since the brand-choice is static, we can write down the expected utility (i.e. Emax) conditional on size:

\[ \omega_t(x) = E \left( \max_j \beta a_{jxt} - \alpha p_{jxt} + \xi_{jxt} + \epsilon_{jt} \right) \]

\[ = \ln \left( \sum_{j=1}^{J} \exp (\beta a_{jxt} - \alpha p_{jxt} + \xi_{jxt}) \right) \]

• McFadden (1978) calls \( \omega_t(x) \) the “inclusive values”. It summarizes the information contained in the vector of prices and characteristics available in period \( t \), very much like a quality-adjusted price index.

• Dynamic programming problem re-defined:

\[ V(i_t, v_t, \omega_t, \epsilon_t) = \max_{c,x} \gamma \ln (c - v_t) - C(i_{t+1}) + \omega_t(x) + \epsilon_{xt} \]

\[ + \beta E \left[ V(i_{t+1}, v_{t+1}, \omega_{t+1}, \epsilon_{t+1}) | i_t, v_t, \omega_t, \epsilon_t, x, c \right] \]

• To be able to define the state-vector solely as \((i_t, v_t, \omega_t, \epsilon_t)\), we need to impose an addition assumption on the transition process of \( \omega_t \):

\[ \Pr(\omega_{t+1} | a_t, p_t, \xi_t) = \Pr(\omega_{t+1} | \omega_t) \]

This assumption as been called the “Inclusive-value sufficiency” assumption (Gowrisankaran and Rysman 2012).
**Step 2:** Estimate the Markov process for $\omega_t(x)$

<table>
<thead>
<tr>
<th></th>
<th>$\omega_{2t}$</th>
<th>$\omega_{4t}$</th>
<th>$\omega_{2t}$</th>
<th>$\omega_{4t}$</th>
<th>$\omega_{2t}$</th>
<th>$\omega_{4t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{1,t-1}$</td>
<td>0.003</td>
<td>-0.014</td>
<td>0.005</td>
<td>0.014</td>
<td>-0.023</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\omega_{2,t-1}$</td>
<td>0.413</td>
<td>0.033</td>
<td>0.295</td>
<td>0.025</td>
<td>0.575</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.007)</td>
<td>(0.013)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>$\omega_{3,t-1}$</td>
<td>0.003</td>
<td>-0.034</td>
<td>0.041</td>
<td>-0.006</td>
<td>0.027</td>
<td>-0.072</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.020)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>$\omega_{4,t-1}$</td>
<td>0.029</td>
<td>0.249</td>
<td>0.026</td>
<td>0.236</td>
<td>-0.018</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.017)</td>
<td>(0.020)</td>
<td>(0.016)</td>
</tr>
</tbody>
</table>

| $\sum_{t=2}^5 \omega_{1,t-\tau}$ | -0.003         | -0.012        | -0.008        | -0.003        |
|                                  | (0.005)       | (0.004)       | (0.006)       | (0.005)       |
| $\sum_{t=2}^5 \omega_{2,t-\tau}$ | 0.089         | 0.006         | 0.073         | -0.004        |
|                                  | (0.003)       | (0.002)       | (0.005)       | (0.004)       |
| $\sum_{t=2}^5 \omega_{3,t-\tau}$ | -0.008        | -0.009        | -0.004        | -0.016        |
|                                  | (0.003)       | (0.003)       | (0.008)       | (0.006)       |
| $\sum_{t=2}^5 \omega_{4,t-\tau}$ | -0.013        | 0.018         | -0.008        | 0.056         |
|                                  | (0.003)       | (0.003)       | (0.007)       | (0.005)       |

*Each column represents the regression of the inclusive value for a size (32, 64, 96, and 128 ounces, respectively) on lagged values of all sizes. The inclusive values were computed using the results in column (x) of Table IV. The four left columns impose the same process for each household type; the four right columns allow for a different process for each type. Reported results are only for households of type 3, that is, households in market 1 with large families. Results for other types are available from the authors.*
• **Step 3:** Dynamic decisions

− Conditional on buying size $x$, the optimal consumption path solves the following continuous DP problem:

$$v(x|s_t) = \max_{c \geq 0} \gamma \ln(c - v_t) - C(i_t - c + x) + \beta E \left[ V(s_{t+1} | s_t, x, c) \right]$$

and the value function takes the familiar form:

$$V(s_t) = \ln \left( \sum_{x=0}^{4} \exp(v(x|s_t)) \right)$$

− The CCP used in the estimation of the dynamic purchasing decision model is:

$$\Pr(x_{it}|s_{it}) = \frac{\exp(v(x_{it}|s_{it}))}{\sum_{x'} \exp(v(x'|s_{it}))}$$

− Remaining problem: $i_t$ is unobserved.

− Solution: Integrate out the unobserved initial inventory levels. How?

  * Let $i_{i0} = 0$.
  * Simulate the model for the first $T_0$ weeks of the data
  * Record inventory level at $T_0$: $i_{T0}$ (i.e. initial state variable)
  * Evaluate the likelihood from weeks $T_0$ to $T$, as if inventory levels were observed
  * Repeat the process $S$ times, and average the likelihood contribution of individual $i$.

− Implicit assumption: Stationary process at time $T_0$. 

40
### TABLE VI
**Third Step: Estimates of Dynamic Parameters**

<table>
<thead>
<tr>
<th>Household Type:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban Market</td>
<td>Suburban Market</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Size:</td>
<td>1-2</td>
<td>3-4</td>
<td>5+</td>
<td>1-2</td>
<td>3-4</td>
<td>5+</td>
</tr>
<tr>
<td>Cost of inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>9.24</td>
<td>6.49</td>
<td>21.96</td>
<td>4.24</td>
<td>4.13</td>
<td>11.75</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.09)</td>
<td>(0.01)</td>
<td>(0.17)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>Quadratic</td>
<td>-3.82</td>
<td>1.80</td>
<td>-35.86</td>
<td>-8.20</td>
<td>-6.14</td>
<td>-0.73</td>
</tr>
<tr>
<td></td>
<td>(29.8)</td>
<td>(1.77)</td>
<td>(0.19)</td>
<td>(0.03)</td>
<td>(1.69)</td>
<td>(1.53)</td>
</tr>
<tr>
<td>Utility from consumption</td>
<td>1.31</td>
<td>0.75</td>
<td>0.51</td>
<td>0.08</td>
<td>0.92</td>
<td>3.80</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.09)</td>
<td>(0.21)</td>
<td>(0.03)</td>
<td>(0.18)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>365.6</td>
<td>926.8</td>
<td>1,530.1</td>
<td>1,037.1</td>
<td>543.6</td>
<td>1,086.1</td>
</tr>
</tbody>
</table>

*aAsymptotic standard errors are shown in parentheses. Also included are size fixed effects, which are allowed to vary by household type.

### TABLE VII
**Long-Run Own- and Cross-Price Elasticities**

<table>
<thead>
<tr>
<th>Brand</th>
<th>Size (oz.)</th>
<th>Allb</th>
<th>Wisk</th>
<th>Surf</th>
<th>Cheer</th>
<th>Tide</th>
<th>Private Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>32</td>
<td>0.418</td>
<td>0.129</td>
<td>0.041</td>
<td>0.053</td>
<td>0.131</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.482</td>
<td>0.093</td>
<td>0.052</td>
<td>0.033</td>
<td>0.085</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.725</td>
<td>0.092</td>
<td>0.036</td>
<td>0.035</td>
<td>0.100</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>-2.556</td>
<td>0.154</td>
<td>0.088</td>
<td>0.059</td>
<td>0.115</td>
<td>0.007</td>
</tr>
<tr>
<td>Wisk</td>
<td>32</td>
<td>0.088</td>
<td>0.702</td>
<td>0.046</td>
<td>0.012</td>
<td>0.143</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.078</td>
<td>0.620</td>
<td>0.045</td>
<td>0.014</td>
<td>0.116</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.066</td>
<td>0.725</td>
<td>0.051</td>
<td>0.022</td>
<td>0.135</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.126</td>
<td>-2.916</td>
<td>0.083</td>
<td>0.026</td>
<td>0.147</td>
<td>0.005</td>
</tr>
<tr>
<td>Surf</td>
<td>32</td>
<td>0.047</td>
<td>0.061</td>
<td>0.977</td>
<td>0.024</td>
<td>0.369</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.146</td>
<td>0.086</td>
<td>0.905</td>
<td>0.023</td>
<td>0.158</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.160</td>
<td>0.101</td>
<td>0.915</td>
<td>0.016</td>
<td>0.214</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.202</td>
<td>0.149</td>
<td>-3.447</td>
<td>0.039</td>
<td>0.229</td>
<td>0.008</td>
</tr>
<tr>
<td>Cheer</td>
<td>64</td>
<td>0.168</td>
<td>0.049</td>
<td>0.027</td>
<td>0.831</td>
<td>0.293</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.167</td>
<td>0.015</td>
<td>0.008</td>
<td>0.982</td>
<td>0.470</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.250</td>
<td>0.090</td>
<td>0.058</td>
<td>4.341</td>
<td>0.456</td>
<td>0.003</td>
</tr>
<tr>
<td>Tide</td>
<td>32</td>
<td>0.071</td>
<td>0.085</td>
<td>0.050</td>
<td>0.022</td>
<td>1.007</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.048</td>
<td>0.055</td>
<td>0.024</td>
<td>0.025</td>
<td>0.924</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.045</td>
<td>0.063</td>
<td>0.016</td>
<td>0.026</td>
<td>1.086</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.072</td>
<td>0.093</td>
<td>0.039</td>
<td>0.045</td>
<td>-2.683</td>
<td>0.001</td>
</tr>
<tr>
<td>Solo</td>
<td>64</td>
<td>0.066</td>
<td>0.070</td>
<td>0.027</td>
<td>0.021</td>
<td>0.150</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.219</td>
<td>0.032</td>
<td>0.023</td>
<td>0.033</td>
<td>0.075</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.127</td>
<td>0.125</td>
<td>0.060</td>
<td>0.043</td>
<td>0.302</td>
<td>0.001</td>
</tr>
<tr>
<td>Era</td>
<td>32</td>
<td>0.035</td>
<td>0.155</td>
<td>0.039</td>
<td>0.022</td>
<td>0.425</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.030</td>
<td>0.103</td>
<td>0.039</td>
<td>0.018</td>
<td>0.304</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>0.035</td>
<td>0.168</td>
<td>0.033</td>
<td>0.027</td>
<td>0.352</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.054</td>
<td>0.192</td>
<td>0.061</td>
<td>0.029</td>
<td>0.513</td>
<td>0.014</td>
</tr>
<tr>
<td>Private</td>
<td>64</td>
<td>0.123</td>
<td>0.119</td>
<td>0.066</td>
<td>0.039</td>
<td>0.081</td>
<td>0.248</td>
</tr>
<tr>
<td>label</td>
<td>128</td>
<td>0.174</td>
<td>0.266</td>
<td>0.100</td>
<td>0.019</td>
<td>0.072</td>
<td>-2.682</td>
</tr>
<tr>
<td>No purchase</td>
<td>0.007</td>
<td>0.002</td>
<td>0.004</td>
<td>0.000</td>
<td>0.013</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

*aCell entries j and j , where i indexes row and j indexes column, give the percent change in market share of brand i with a 1 percent change in the price of j. All columns are for a 128 oz. product, the most popular size. The results are based on Tables IV–VI.

*bNote that "All" is the name of a detergent produced by Unilever.
- Two differences between static and dynamic model:
  1. Static model implies larger price coefficient,
  2. and ignores the inventory problem
- Both lead to a larger **own** price elasticity with the static model, than the long-run own price elasticity with the dynamic model.
- Point two leads to a **lower** cross price elasticity (with the static model). Why?
  - In the data the response to sales is mostly coming from people going from not buying to buying the brand on sale.
  - This leads to predict small cross price elasticities with respect to other products, and large cross price elasticity with respect to the outside good.
  - Or more mechanically, in the static model, the choice probability are all relatively small since consumers buy detergent infrequently. The logic cross-partials are equal to the product of the two choice-probabilities: small cross-elasticities.
  - The dynamic model rationalize this phenomenon high unobserved inventories, and conditional choice probabilities (conditional on purchasing). Everybody needs detergent...

<table>
<thead>
<tr>
<th>TABLE VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AVERAGE RATIOS OF ELASTICITIES COMPUTED FROM A STATIC MODEL TO LONG-RUN ELASTICITIES COMPUTED FROM THE DYNAMIC MODEL</strong></td>
</tr>
<tr>
<td>Brand</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>All</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Wisk</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Surf</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cheer</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tide</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Solo</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Era</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Private</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>No purchase</td>
</tr>
</tbody>
</table>

1. Call entries i and j, where i denotes row and j denotes column, give the ratio of the (short-run) elasticities computed from a static model divided by the long-run elasticities computed from the dynamic model. The elasticities for both models are the percent change in market share of brand i with a 1 percent change in the price of j. The static model is identical to the model estimated in the first step, except that brands of all sizes are included as well as a no-purchase decision, not just products of the same size as the chosen option. The results from the dynamic model are based on the results presented in Tables IV–VI.
2. Note that “All” is the sum of a detergent produced by Unilever.
Demand for experience goods and brand loyalty

- **Introduction**: Measuring state dependence in consumer choice behavior. In marketing and economics it is frequently observed that demand for some products exhibit time dependence:

\[
\Pr(d_{ijt} = 1|d_{ijt} = 1) > \Pr(d_{ijt} = 1|d_{ijt} = 0).
\]

- **Examples**:
  - Switching cost
  - Brand loyalty
  - Persistent unobserved heterogeneity

- Many papers have been written to empirically distinguish between true state-dependence and unobserved heterogeneity (e.g. Keane (1997)):

\[
U_{ijt} = X_{ijt}\beta + \lambda H_{ijt} + \epsilon_{ijt}
\]

where \(\epsilon_{ijt} = \rho\epsilon_{ijt-1} + \nu_{ijt}\)

\[
H_{ijt} = \sum_{\tau=t_0}^{t-1} g(\tau)d_{ij\tau}
\]

- **Empirical challenges**:
  - Initial condition problem (i.e. \(d_{t_0}\) is endogenous and/or unobserved).
  - The presence of persistent unobserved heterogeneity can generate “spurious” state-dependence.
  - Multi-dimension integration when \(\epsilon_{ijt}\) are correlated across options and time:

\[
\Pr(d_i|d_{t_0}) = \Pr(U_{ijt} > U_{ikt}, \forall k \neq j, d_{ijt} = 1, t = 1...T) = \Pr(\epsilon_{ijt} - \epsilon_{ikt} > -(X_{ijt} - X_{ikt})\beta - (H_{ijt} - H_{ikt})\lambda, \\
\forall k \neq j, d_{ijt} = 1, t = 1...T),
\]

i.e. dimension is \(T \times (J - 1)\). Cannot be evaluated with standard methods. We must use simulation.
GHK Simulator

- Simpler cross-sectional example with 4 choices:
  \[ U_{ij} = X_{ij} \beta + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \Omega) \]  
  (5)

Then the probability of choosing option 4 is:

\[
\Pr(d_{i4} = 1) = \Pr(\epsilon_{ij} - \epsilon_{i4} < -(X_{ij} - X_{i4}) \beta, \forall k \neq 4) \\
= \int_{A_1 \cap A_2 \cap A_3} dF(\epsilon_{i1} - \epsilon_{i4}, \epsilon_{i1} - \epsilon_{i3}, \epsilon_{i3} - \epsilon_{i4})
\]

where \( A_j = \{\epsilon_{ij} - \epsilon_{i4} < -(X_{ij} - X_{i4}) \beta\} \)

Let redefine the variables to compute the probability of choosing option 4:

\[
\nu_{ij} = \epsilon_{ij} - \epsilon_{i4} \\
X^*_{ij} = X_{ij} - X_{i4} \\
\nu \sim N(0, \Sigma), \quad \Sigma_{3 \times 3} = C'C \\
\nu = C'\eta, \quad \eta \sim N(0, I).
\]

- Standard Monte-Carlo integration (i.e. Accept-reject):

1. Draw \( M \) vectors \( \eta^m \sim (0, I) \),
2. Keep draw \( m \) if \( C'\eta^m \in A_1 \cap A_2 \cap A_3 \). Otherwise reject.
3. Compute simulated choice probability:

\[
\hat{\Pr}(d_{i4}) = \frac{\text{Number accepted draws}}{M}
\]

(6)

Problems and limitations:

- Non-smooth simulator (i.e. cannot use gradient methods)
- Require a high number of draws to avoid \( Pr(d_{i4}) = 0 \).
If the dimension of integration is large: infeasible (e.g. Panel data). We can alleviate the non-smooth problem by smoothing the Accept/Reject probability:

\[
\hat{\Pr}(d_{i4}) = \frac{1}{4} \sum_m \frac{1}{1 + \sum_k \exp\left(( -X_{ik}^*\beta - \nu^m_{ij})/\rho \right)}
\]

...still very bias if \( M \) is too small.

- The GHK simulator avoids the main problems of the standard MC-AR method by drawing only from the accepted region.

Recall that:

\[
C = \begin{pmatrix}
c_{11} & c_{12} & c_{13} \\
0 & c_{22} & c_{23} \\
0 & 0 & c_{33}
\end{pmatrix}
\]

In order to compute \( \hat{\Pr}(d_{i4} = 1) \) we proceed sequentially:

1. Draw \( \nu^m_{i1} \):
   
   Compute \( \Phi_{i1} = \Pr(\nu_{i1} < -X_{i1}^*\beta) \).
   
   Draw \( \eta^m_{i1} \) from a truncated normal:
   
   \( \eta \sim \Phi(-X_{i1}^*/c_{11}) \)
   
   How?
   
   (a) Draw \( \lambda_i \sim U[0, 1] \)
   
   (b) Set \( \lambda_{i1} = \lambda_i \Phi_{i1} \)
   
   (c) Set \( \eta^m_{i1} = \Phi^{-1}(\lambda_{i1}) \)
   
   (d) Finally \( \nu_{i1} = c_{11}\eta^m_{i1} \)

2. Draw \( \nu^m_{i2} \) from a truncated normal (conditional on \( \nu^m_{i1} \)):

\[
\nu^m_{i2} = c_{12}\eta^m_{i1} + c_{22}\eta_{i2} < -X_{i2}^*\beta
\]

\[
\eta^m_{i2} \sim \Phi\left(\frac{-X_{i2}^*\beta - c_{12}\eta^m_{i1}}{c_{22}}\right) \equiv \Phi_{i2}
\]

Thus we first draw \( \eta^m_{i2} \) from \( \Phi\left(\frac{-X_{i2}^*\beta - c_{12}\eta^m_{i1}}{c_{22}}\right) \) as before, and compute

\( \nu^m_{i2} = c_{12}\eta^m_{i1} + c_{22}\eta^m_{i2} \).
3. Compute \( \Phi_{i3} \) similarly.

4. Compute \( \hat{\Pr}(d_{i4} = 1) \):

\[
\hat{\Pr}(d_{i4} = 1) = \frac{1}{M} \sum_{m} \Phi\left( \frac{-X_{i1}^*\beta}{c_{11}} \right) \times \Phi\left( \frac{-X_{i2}^*\beta - c_{12}\eta_{m}^{i1}}{c_{22}} \right) \\
\times \Phi\left( \frac{-X_{i3}^*\beta - c_{12}\eta_{m}^{i1} - c_{23}\eta_{m}^{i2}}{c_{33}} \right)
\]

- Advantages of the GHK:
  - Highly accurate even with high dimension integrals
  - Differentiable
  - Require fewer draws

- In order to applied to panel data with AR(1) correlation in the \( \epsilon_{ij} \) and arbitrary correlation across options we need to simulate the probability of observing a sequence of choices \( \{j_{it}\}_{t=1}^{T} \).

- The Algorithm is the same... just longer (i.e. there are \((J - 1)T\) sequential draws to make for each \(m\) and \(i\)). To see this, let redefine the variables in the following way:

\[
\tilde{U}_{kt} = U_{kt} - U_{j_{it}} \\
\tilde{\epsilon}_{kt} = \epsilon_{kt} - \epsilon_{j_{it}}
\]

Where, \( \tilde{\epsilon} \sim N(0, \tilde{\Sigma}) \),

Where \( \tilde{\Sigma} = \tilde{C} \tilde{C}' \) is a \((J - 1)T \times (J - 1)T\) covariance matrix appropriately transformed to reflect the vector of choices \( j_{it}, t = 1...T \).

- With this transformation, \( i \rightarrow j_{it} \) if \( \tilde{U}_{kt} \leq 0, \forall k \) and \( \tilde{\epsilon}_{i} = \tilde{C}'\eta_{i} \).

- Then to compute \( \hat{\Pr}(j_{i}) \):

Period 1: Draw \( \tilde{\epsilon}_{i1}^{m} \):

1: Draw \( \eta_{i11}^{m} \) from a truncated normal s.th:

\[ \tilde{U}_{i11}(\eta_{i11}^{m}) < 0. \]
2: Draw $\eta_{i21}^m$ from a truncated normal s.th:

$$\tilde{U}_{i21}(\eta_{i11}^m, \eta_{i21}^m) < 0.$$

... 

$j_{i1}$: Skip $\nu_{j_{i11}}^m$ 

... 

Period $t$: Draw $\tilde{\epsilon}_{it}^m$:

1: Draw $\eta_{i1t}^m$ from a truncated normal s.th:

$$\tilde{U}_{i11}(\eta_{i11}^m, \ldots, \eta_{iJt}^m, \eta_{i1t}^m) < 0.$$

...

$j_{i1} - 1$: Draw $\eta_{ijit-1t}^m$ from a truncated normal s.th:

$$\tilde{U}_{ij1-1t}(\eta_{i11}^m, \ldots, \eta_{iJt-1}, \eta_{i1t}^m, \ldots, \eta_{ijit-1t}^m) < 0).$$

$j_{i1}$: Skip $\eta_{ijit}^m$ 

...

Finally compute $\hat{\Pr}(j_i)$ by taking the product of each component and averaging over $m$.

- Keane (1997) estimates many different specifications of equation 4 using SMS with GHK and finds strong evidences of “true” state-dependence for Ketchup.

- Erdem and Keane (1996), Ackerberg (2003), and Crawford and Shum (2005) all estimate structural models of quality uncertainty that generate endogenously this type of brand loyalty or switching costs.

- Estimate a dynamic bayesian learning model of demand for detergent using the A.C. Nielsen \textbf{public} scanner data set (i.e. \(\sim\) 3000 households over 3 years). The data-sets for detergent, ketchup, margarine and canned soup are available at:

  \url{http://research.chicagogsbe.edu/marketing/databases/erim/index.aspx}

- Why?

  - Market with frequent brand introduction.
  - Reasonable to assume that consumers learn about the quality of the product \textbf{only} through experience and advertising.
  - Provide and economic interpretation for brand loyalty: If consumers are risk averse experiencing “too” frequently is costly (i.e. endogenous switching cost).
  - Measure the information content of advertising messages (see also Ackerberg (2001) and Ackerberg (2003)).
Model

- Finite Horizon DP problem:

\[
V_j(I(t), d_j = 1) = \begin{cases} 
E[U_j(t)|I(t)] + e_{jt} + \beta E[V(I(t+1)|I(t), d_j(t)] & \text{if } t < T \\
E[U_j(T)|I(T)] + e_{jt} & \text{else}
\end{cases}
\]

\[
V(I(t)) = \max_{d_j, j = 1..J} V_j(I(t), d_j)
\]

Where the expectation is taken over the evolution of the information set \(I(t)\) and the idiosyncratic shock \(e_{jt}\).

- Components of the expected utility:

1. Attribute of good \(j\) if “experienced” at \(t\):

\[
A^{E}_{jt} = A_j + \delta_{jt}.
\]

2. Utility after realization of \(\delta_{jt}\) and \(e_{jt}\):

\[
U_{jt} = -\omega_p P_{jt} + \omega_A A^{E}_{jt} - \omega_A r A^{E^2}_{jt} + e_{jt}
\]

Where \(r\) = risk aversion coefficient

\(e_{jt}\) = “logit” utility shock

\(P_{jt}\) = Price of \(j\) (stochastic)

3. Expected utility:

\[
E[U_{jt}|I(t)] = -\omega_p P_{jt} + \omega_A E[A^{E}_{jt}|I(t)] - \omega_A r E[A^{E^2}_{jt}|I(t)] - \omega_A r E[(A^{E}_{jt} - E[A^{E}_{jt}|I(t)])^2|I(t)] + e_{jt}
\]

- Outside options:

\[
E[U_{0t}|I(t)] = \Phi_0 + \Psi_0 t + e_{0t}
\]

\[
E[U_{NPt}|I(t)] = \Phi_{NP} + \Psi_{NP} t + e_{NPt}
\]
• Signals and components of the information set:

1. Experience:
   \[ A_{jt}^E = A_j + \delta_{jt}, \quad \delta_{jt} \sim N(0, \sigma_\delta^2) \]
   * The experience signals are unbiased: \( \delta \) is mean zero.

2. Priors on \( A_j \):
   \[ A_j \sim N(A, \sigma_\nu(0)^2). \]

3. Advertising message (with probability \( p_j^S \) estimated from the data):
   \[ S_{jt} = A_j + \eta_{jt}, \quad \eta_{jt} \sim N(0, \sigma_\eta^2), \]
   * The information content of advertising messages is measured by \( \sigma_\eta^2 \)
   * Advertising is exogenous and strictly informative: \( \eta_{jt} \) is mean zero.

• Bayesian Updating: Update expectation about good \( j \)'s attribute
   \[
   E[A_j|I(t)] = E[A_j|I(t-1)] + d_{jt} \beta_{1jt} [A_{jt}^E - E[A_{jt}^E|I(t-1)]] \\
   + ad_{jt} \beta_{2jt} [S_{jt} - E[S_{jt}|I(t-1)]]
   \]
   Where the updating weights are given by:
   \[
   \beta_{1jt} = \frac{\sigma_{\nu_j}^2(t)}{\sigma_{\nu_j}^2(t) + \sigma_\delta^2} \quad \& \quad \beta_{2jt} = \frac{\sigma_{\nu_j}^2(t)}{\sigma_{\nu_j}^2(t) + \sigma_\eta^2}
   \]

• From the econometrician point of view (since \( A_j \) is a parameter), we can rewrite the problem in terms of expectation errors \( \nu_j(t) = E[A_{jt}^E|I(t)] - A_j \). This generates a first-order markov process payoff in relevant state variables:
   \[
   \nu_j(t) = \nu_j(t-1) + d_{jt} \beta_{1jt} [-\nu_j(t-1) + \delta_{jt}] \\
   + ad_{jt} \beta_{2jt} [-\nu_j(t-1) + \eta_{jt}]
   \]
   With \( \nu_j(0) = A - A_j, \forall j. \)
Similarly, the precision of signals is updated using the following markov process:

\[
\sigma_{\nu_j}(t) = \left[ \frac{1}{\sigma_{\nu}(0)} + \frac{\sum_{s \leq t} d_{js}}{\sigma_{\delta}^2} + \frac{\sum_{s \leq t} a d_{js}}{\sigma_{\nu}^2} \right]^{-1}
\]  

(9)

• Timing:

\[ t = 0 \]  * Purchasing decision based on \( \nu_j(0) \).
  * New signals: \( \delta_{j0} \) (if \( d_{j0} = 1 \)) and \( \eta_{j0} \) (if \( a d_{j0} = 1 \)).
  * Update \( \nu_j(1) \) according to equation 8

\[ t = 1 \]  * Purchasing decision based on \( \nu_j(1) \).
  * New signals: \( \delta_{j1} \) (if \( d_{j1} = 1 \)) and \( \eta_{jt} \) (if \( a d_{j1} = 1 \)).
  * Update \( \nu_j(2) \) according to equation 8

...  

• Therefore at any period \( t \) the payoff relevant state variables are:

\[
I(t) = \left\{ \sum_{s \leq t-1} d_{js}, \sum_{s \leq t-1} a d_{js}, \nu_{jt} \right\}_{j=1...J}
\]

Discrete \hspace{2cm} Continuous

• Solution/Estimation: Nested fixed-point estimation algorithm where the DP is solved by backward induction.

• Challenges:

  – Size of the state-space: Impossible to solve exactly.
  – Choice probabilities:

\[
Pr_j(I(t)) = \int \frac{\exp \left( EU_j(I(t)) + \beta E[V(I(t + 1), t + 1)] \right)}{\sum_k \exp \left( EU_j(I(t)) + \beta E[V(I(t + 1), t + 1)] \right)} f(\nu) d\nu
\]

Integration is complicated by the fact that \( \nu_{jt} \) is serially correlated: Must integrate the sequence of past \( \nu_{js} \)'s using simulation method: simulate \( M \) sequences of \( \nu_{jt} \), just like in Hendel and Nevo (2006a).
– Initial condition problem: Do not observe the initial level of purchasing history and attribute expectation (problem if the products are not newly introduced as in Ackerberg (2003)).

**Solution:** Set $I(0) = \{0, 0, \nu_{j0}\}_{j=1,\ldots,J}$, and simulate the model for the first two years of the data. Use the last two years for the estimation (i.e. $T = 100$ weeks).
Solution Method: Keane and Wolpin (1994)

- **General Idea:** Solve the value function exactly only at a subset $I^*(t)$ of the states and interpolate between them using Least-squares to compute $EV(I(t))$ at $I(t) \notin I^*$.

- Backward induction algorithm for a fix grid $I^*$:

  **$T$:** 1. Calculate $EV_T(I(T))$ for all $I(T) \in I^*$:

  $$EV_T(I(T)) = \int \left\{ \max_j EU_{jT}(I(T)) + e_{jT} \right\} dF(e)$$

  $$= \log \left( \sum_j \exp \left( EU_{jT}(I(T)) \right) \right)$$

  2. Run the following regression:

  $$EV_T(I(T)) = G(I(T))\theta^T + \nu = \hat{EV}_T(I(T)) + u,$$

  where $G(I(T))$ is vector containing flexible transformations of the state variables, where $u$ is a regression error.

  **Note:** For the approximation to work, the $R^2$ of the regressions must be very high. Alternative methods exist to improve the quality of the interpolation (e.g. kernels, local polynomials, etc).

  **$T-1$:** 1. Draw $M$ random variables: $\{\delta_1^m, \ldots, \delta_J^m, \eta_1^m, \ldots, \eta_J^m, ad_1^m, \ldots, ad_J^m\}$.

  2. For each state $I(T-1) \in I^*$ compute the expected value of choosing brand $j$ in $T-1$:

  $$E[V_T(I(T))|I(T-1), d_{jT-1} = 1] = \frac{1}{M} \sum_m \hat{EV}_T(I^m(T))$$

  Where $I^m(T)$ is the state corresponding to the $m$th draw and $d_{jT-1} = 1$. If $I^m(T) \in I^*$ use the exact solution, otherwise use $G(I^m(T))\theta^T$. 

53
3. For each $I(T - 1) \in \mathcal{I}^*$ calculate $V(I(T - 1))$:

$$EV_{T-1}(I(T - 1)) = \log \left( \sum_j \exp \left( EU_{jT-1}(I(T - 1)) + \beta E[V_T(I(T))|I(T - 1), d_{jT-1} = 1] \right) \right)$$

4. Run the following regression:

$$EV_{T-1}(I(T - 1)) = G(I(T - 1))\theta^{T-1} + \nu,$$

... Repeat steps 1-4 for the remaining periods $t = T - 2, ..., 0$.

- **Note**:
  - In Erdem and Keane (1996) $G(I(t))$ includes the expected attributed level of each brand and the perception error variances.
  - To estimate the model, the model needs to be solve using the Interpolation/Simulation algorithm for each parameter values.
  - The choice probabilities are computed by simulating a sequence of states for each households.
Results

- Risk-aversion coefficient is large and negative: Important switching cost of experimenting.

- The variance of the advertising signal is much much larger than the variance of the experience signal: Consumers don’t get much from TV ads!

- Initial priors are very precise: Little uncertainty in this market.

- Structural model fits better...

<table>
<thead>
<tr>
<th>Table 3 Structural Model Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>price coefficient ($-w_a$)</td>
</tr>
<tr>
<td>utility weight ($w_a$)</td>
</tr>
<tr>
<td>risk coefficient ($r$)</td>
</tr>
<tr>
<td>initial variance ($\sigma^2(t)$)</td>
</tr>
<tr>
<td>mean attribute levels ($A_i$):</td>
</tr>
<tr>
<td>$A_{\text{Cash}}$</td>
</tr>
<tr>
<td>$A_{\text{Cheer}}$</td>
</tr>
<tr>
<td>$A_{\text{Solo}}$</td>
</tr>
<tr>
<td>$A_{\text{Surf}}$</td>
</tr>
<tr>
<td>$A_{\text{Fam}}$</td>
</tr>
<tr>
<td>$A_{\text{Wish}}$</td>
</tr>
<tr>
<td>$A_{\text{Hole}}$</td>
</tr>
<tr>
<td>“Other Brands” intercept ($\Phi_0$)</td>
</tr>
<tr>
<td>“Other Brands” time trend ($\Psi_0$)</td>
</tr>
<tr>
<td>“No Purchase” intercept ($\Phi_{\text{NP}}$)</td>
</tr>
<tr>
<td>“No Purchase” time trend ($\Psi_{\text{NP}}$)</td>
</tr>
<tr>
<td>experience variability ($\sigma_0$)</td>
</tr>
<tr>
<td>advertising variability ($\sigma_1$)</td>
</tr>
</tbody>
</table>

$^1$ – LL = 7312.09 AIC = 7324.09 BIC = 7384.49

$^2$ – LL = 7306.05 AIC = 7322.05 BIC = 7378.45
“Across several treatment lengths and spell transitions, there is a marked decreasing trend in the switching probability at the very beginning of treatment.”

Two forces explain these switching probabilities in the bayesian learning model:

- Initial experimentation and risk aversion.
- Forward looking behavior: Incentive for patients (or doctors) to acquire more information by experimenting.
Model

• A treatment is characterized by two match values (contrary to only one in Erdem and Keane (1996)):
  
  $- \mu_{jn} \Rightarrow$ Symptomatic or side-effects (enters the utility directly)
  $- \nu_{jn} \Rightarrow$ Curative properties (enters the recovery probability).

• Signals about the match values if $j$ use drug $n$ at $t$:

$$x_{jnt} \sim N(\mu_{jn}, \sigma^2_n)$$
$$y_{jnt} \sim N(\nu_{jn}, \tau^2_n)$$

• Initial priors about the match values:

$$\mu_{jnt} \sim N(\bar{\mu}_{nk}, \bar{\sigma}^2_n)$$
$$\nu_{jnt} \sim N(\bar{\nu}_{nk}, \bar{\tau}^2_n)$$

Where $k = 1...4$ indexes the severity type of patients (learned perfectly by the initial diagnostic).

• Expected Utility (CARA):

$$u(x_{jnt}, p_n, \epsilon_{jnt}) = -\exp(-rx_{jnt}) - \alpha p_n + \epsilon_{jnt}$$
$$\widetilde{EU}(\mu_{jn}(t), \nu_{jn}(t), p_n, \epsilon_{jnt}) = -\exp \left( -r \mu_{jn}(t) + 1/2r^2(\sigma^2_n + V_{jn}(t)) \right)$$
$$-\alpha p_n + \epsilon_{jnt}$$
$$= \text{EU} (\mu_{jn}(t), V_{jn}(t), p_n) + \epsilon_{jnt}$$

• Recovery probability follow a markov process:

$$h_j(t) = \frac{\left( \frac{h_j(t-1)}{1-h_j(t-1)} \right) + d_{jnt}y_{jnt}}{1 + \left( \frac{h_j(t-1)}{1-h_j(t-1)} \right) + d_{jnt}y_{jnt}}$$
• Updating rule for the beliefs regarding the symptomatic and curative match value $\mu_{jn}(t + 1)$ and $\nu_{jn}(t + 1)$ by equation (7) and (8) (same as in Erdem and Keane).

• State space:

$$s_{jt} = \left\{ \mu_{jn}(t), \nu_{jn}(t), l_{jn}(t), h_j(t) \right\}_{n=1...5}$$

Where $l_{jn}(t) = \sum_{s<t} d_{jnt}$.

• **Value Function:** Infinite horizon problem with absorbing state (i.e. recovery)

$$V(s) = \int \max_n EU(s) + \epsilon_n + \beta E[(1 - h(s'))V(s')|d_n = 1, s]$$

$$= \log \left[ \sum_n \exp (EU(s) + \beta E[(1 - h(s'))V(s')|d_n = 1, s]) \right]$$

• **Solution method:** Value function iteration with interpolation and simulation (i.e. Keane and Wolpin (1994))

1. Define a discrete grid $S^* \in S$.
2. For each state $s \in S^*$ make an initial guess at the value function $V^0(s)$.
3. Run regression:

$$V^0(s) = G(s)'\theta^0 + u_s$$

4. Draw $M$ random signals $\{x_{jn}^m, y_{jn}^m\}$
5. Compute the expected value of choosing drug $n$ for each $s \in S^*$:

$$E[V(s|d_n = 1, s)] = \frac{1}{M} \sum_m (1 - h(s^m))V^0(s^m)$$

Where $s^m$ is state corresponding to the random draw $m$ and drug $n$ being chosen, and $V^0(s^m)$ is evaluated with the interpolation equation if necessary.
6. Update the value function for each \( s \in S^* \):

\[
V^1(s) = \log \left[ \sum_n \exp \left( EU(s) + \beta E \left[ V(s|d_n = 1, s) \right] \right) \right]
\]

7. Repeat step 3-6 until convergence of the value function at the grid points.

**Results**

- Large risk aversion coefficient: Important switching cost
- Types are horizontally differentiated
- Policy experiments:
  - Concentration increases in the level of uncertainty (i.e. smaller switching costs)
  - The “pooling” of types (i.e. poor initial diagnostic) decreases concentration (i.e. products become less differentiated).
Table III
Dynamic Model: Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illness heterogeneity distribution</td>
<td>Recovery Probability</td>
<td>Type Probability</td>
</tr>
<tr>
<td>( \theta_1 ) (Type 1)</td>
<td>0.433</td>
<td>0.593</td>
</tr>
<tr>
<td>( \theta_2 ) (Type 2)</td>
<td>0.127</td>
<td>0.335</td>
</tr>
<tr>
<td>( \theta_3 ) (Type 3)</td>
<td>0.199</td>
<td>0.043</td>
</tr>
<tr>
<td>( \theta_4 ) (Type 4)</td>
<td>0.432</td>
<td>0.029</td>
</tr>
<tr>
<td>Means, symptom match values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>0.927</td>
<td>1.195</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>0.928</td>
<td>0.428</td>
</tr>
<tr>
<td>( \mu_3 )</td>
<td>0.481</td>
<td>0.483</td>
</tr>
<tr>
<td>( \mu_4 )</td>
<td>0.325</td>
<td>0.145</td>
</tr>
<tr>
<td>( \mu_5 )</td>
<td>0.451</td>
<td>0.174</td>
</tr>
<tr>
<td>Means, curative match values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. dev., symptom match values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma )</td>
<td>1.574</td>
<td>0.006</td>
</tr>
<tr>
<td>Std. dev., symptom signals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>0.998</td>
<td>0.006</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>1.134</td>
<td>0.006</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>1.375</td>
<td>0.006</td>
</tr>
<tr>
<td>( \sigma_4 )</td>
<td>1.159</td>
<td>0.006</td>
</tr>
<tr>
<td>( \sigma_5 )</td>
<td>0.931</td>
<td>0.006</td>
</tr>
<tr>
<td>Std. dev., curative match values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau )</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Std. dev., curative signals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau )</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>Price coefficient, ( a^d )</td>
<td>1.080</td>
<td>0.991</td>
</tr>
<tr>
<td>Risk-aversion parameter, ( \alpha )</td>
<td>0.990</td>
<td>0.274</td>
</tr>
<tr>
<td>Discount rate, ( \beta )</td>
<td>0.950</td>
<td>Fixed</td>
</tr>
<tr>
<td>Number of observations</td>
<td>34,972</td>
<td></td>
</tr>
<tr>
<td>Number of similar draws</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Log likelihood function</td>
<td>-124,484.34</td>
<td></td>
</tr>
</tbody>
</table>

\( ^d \) Price in thousands of lire.

\( ^b \) Prior symptomatic and curative means are reported for two highest probability types. For prior means corresponding to all types, see Table IV.

\( ^c \) Prior means for drug 2 are reported for the Jul–Dec 1992 period. For prior means corresponding to earlier periods, see Table IV.
Demand for durable goods

Motivation:

• In many settings, the timing of purchase is as important as what you purchase.

• In durable goods markets, this is true because the quality and price of products available at any point in time vary less, than the quality and price of products offer across time.

• Why? Introduction of new products push the technology frontier upwards, and most firms often respond by using dynamic pricing strategy (i.e. discount products that are about to be discontinued).

• Challenges: (i) Micro data on product replacement is rare, and (ii) dynamic “engine” replacement models with product differentiation have very large state space.

• Examples: Gordon (2010), Gowrisankaran and Rysman (2012)
Model: Gowrisankaran and Rysman (2012)

• Indirect utility over product characteristics (omit time):

\[ u_{ij} = \beta_i x_j + \xi_j + \varepsilon_{ij} = \delta_j + \mu_{ij} + \varepsilon_{ij} \]  (10)

• Utility over current product:

\[ u_{i0} = \beta_i x_0 + \varepsilon_{i0} \]  (11)

Note: product 0 is either the outside option (not using a camcorder), or the last camcorder. GR assume that products do not depreciate, and there are no sunk replacement costs.

• Bellman equation:
  
  – Multinomial choice: \( a_i \in A_i = \{0, 1, ..., J\} \). Where \( J \) is the current number of products available.
  
  – Common state-space: \( s_i = \{u_{i0}, x_1, ..., x_J, p_1, ..., p_J\} \)
  
  – Expected Bellman equation:

\[
\bar{V}_i(s_i) = E_{\varepsilon}\left[ \max_{a \in A_i} u_{i,a} - \alpha_i p_{ia} + \varepsilon_{ia} + \beta E_{s'}(\bar{V}_i(s')|s_i, a) \right]
\]

\[
= E_{\varepsilon}\left[ \max \left\{ v_i(0|s_i) + \varepsilon_{i0}, \max_{a \in A\setminus0} v_i(a|s_i) + \varepsilon_{ia} \right\} \right]
\]

where \( p_{ia} = 0 \) if \( a = 0 \), and \( v_i(a|s_i) = u_{ia} - \alpha_i p_{ia} + \beta E_{s'}(\bar{V}_i(s')|s_i, a) \).

• Curse of dimensionality: (i) The number of grid points to approximate \( V_i(s_i) \) grows exponentially with the number of products, and (ii) The calculation of the continuation value involves a \(|s|\) dimension integral (i.e. \( E_{s'}(V|x) \)).
Dimension reduction assumption:

- Recall that if $\varepsilon_{ij}$ is distributed according to a type-1 extreme value distribution with location/spread parameters $(0, 1)$, then $\max_{a \in A \setminus 0} v_i(a|x) + \varepsilon_{ia}$ is also EV1 distributed with location parameter:

$$\omega_i(s_i) = \log \left( \sum_{a \in A \setminus 0} \exp(v_i(a|s_i)) \right)$$

and spread parameter of 1.

- $\omega_i(s_i)$ is the expected discounted value of buying the preferred camcorder available today, instead of holding to the current one (i.e. outside option).

- Therefore, we can rewrite the expected Bellman equation as:

$$\bar{V}_i(s_i) = E_\varepsilon \left[ \max \{v_i(0|s_i) + \varepsilon_{i0}, \omega_i(s_i) + \varepsilon_{i1}\} \right]$$

$$= \log (\exp(v_i(0|s_i)) + \exp(\omega_i(s_i)))$$

- So far, we’re simply re-writing the problem so that it looks more like a sequential decision problem: (i) replace or not replace, and (ii) which new product to buy conditional on replacing. The value of replacing is $\omega_i(s_i)$.

- In order to reduce the dimension of the state space, GR introduce a new assumption:

**Inclusive value sufficiency:** If $\omega_i(s_i) = \omega_i(\tilde{s}_i)$, then $g_\omega(\omega_i(s')|s_i) = g_\omega(\omega_i(s'|\tilde{s}_i))$. Where $g_\omega(x|s_i)$ is the density the option value $\omega$.

- Importantly, this implies that if two different states have the same option value $\omega$, then they also have the same value function.

- In addition, we assume that $\omega_i$ evolves over time according to an AR(1) process:

$$\omega_{i,t+1} = \rho_{i,0} + \rho_{i,1}\omega_{i,t} + \eta_{i,t+1}$$

where $\eta_{i,t}$ is mean zero shock.
• The IVS assumption, means that consumers can use the scalar $\omega$ to forecast the future and evaluate the value relative value of buying now versus waiting.

• New “sufficient” state space: $s_i = (u_{i0}, \omega_i)$.

• New expected value function:

$$V_i(u_{i0}, \omega_i) = E_{\epsilon_i} \left[ \max \{ u_{i0} + \epsilon_{i0} + \beta E_{\omega_i'} [V_i(u_{i0}, \omega_{i'})|\omega_{i'}], \omega_i + \epsilon_{i1} \} \right]$$

$$= \log \left( \exp(u_{i0} + \epsilon_{i0} + \beta E_{\omega_i'} [V_i(u_{i0}, \omega_{i'})|\omega_{i'}]) + \exp(\omega_i) \right)$$

$$= \Gamma_i(u_{i0}, \omega_i|\bar{V}_i)$$

where $\Gamma_i(\bar{V}_i)$ is a contraction mapping from $\bar{V}_i$ to $\bar{V}_i$. 
**Equilibrium conditions:**

1. Rational expectation:

\[ \omega_{i,t+1} = \rho_{i,0} + \rho_{i,1}\omega_{i,t} + \eta_{i,t+1} \]

2. Bellman equation:

\[ \bar{V}_i(u_{i0}, \omega_i) = \log \left( \exp(u_{i0} + \beta E_{\omega'}[\bar{V}_i(u_{i0}, \omega')|\omega]) + \exp(\omega_i) \right) \]

3. Consistency of the inclusive value:

\[ \omega_i = \log \left( \sum_{a \in A \setminus 0} \exp(u_{i,a} + \beta E_{\omega'}[\bar{V}_i(u_{i,a}, \omega'_i)|\omega_i]) \right) \]

4. Aggregate market share:

\[ \hat{s}_{jt} = s_{jt}(\delta, \theta) = \frac{1}{S} \sum_i \sum_{j_0} P_j(u_{i,j_0}, \omega_{it}|\alpha_i, \beta_i)w_{j_0}(\alpha_i, \beta_i) \]

where \( P_j(u_{i,j_0}, \omega_{it}|\beta_i, \alpha_i) \) is the probability of choosing option \( j \) conditional on holding option \( j_0 \) last period, \( w_{j_0}(\alpha_i, \beta_i) \) is the share of consumers of types \((\alpha_i, \beta_i)\) holding option \( j_0 \) in period \( t - 1 \), and \( S \) is the number of simulated consumer.

In order to estimate \( \theta \), we must minimize the following GMM criterium function subject to the equilibrium conditions:

\[
\min_{\theta} \quad (\xi^T \mathbf{Z}) \Omega^{-1} (\xi^T \mathbf{Z})^T \\
\text{s.t.} \quad \delta_{jt} = X_{jt} + \xi_{jt} = s_{jt}^{-1}(\hat{s}, \theta) \\
\omega_{i,t+1} = \rho_{i,0} + \rho_{i,1}\omega_{i,t} + \eta_{i,t+1} \\
\bar{V}_i(u_{i0}, \omega_i) = \Gamma_i(u_{i0}, \omega_i|\bar{V}) = \log \left( \exp(u_{i0} + \beta E_{\omega'}[\bar{V}_i(u_{i0}, \omega')|\omega]) + \exp(\omega_i) \right) \\
\omega_i = \log \left( \sum_{a \in A \setminus 0} \exp(u_{i,a} + \beta E_{\omega'}[\bar{V}_i(u_{i,a}, \omega'_i)|\omega_i]) \right)
\]
**Solution steps:** At every candidate parameter \( \theta \)

1. **Initialization:**
   - Sample time-invariant \( S \) consumer types: \( \beta \)
   - State-space grid: \((u_0, \omega)\)

2. **Outer-loop:** Solve \( \delta_{jt} = s_{jt}^{-1}(s, \theta) \)
   (a) Starting values: \( \delta_{jt}^0 \)
   (b) **Inner-loop:** Solve transition parameters \((\rho_{i,0}, \rho_{i,1})\), and Bellman equation \( \bar{V}_i(u_0, \omega) \)
      i. Initial guesses: \( \bar{V}_i(0, \omega) \) and \((\rho_{i,0}, \rho_{i,1})\), for all \( i = 1, ..., S \)
      ii. Solve inclusive value fixed-point for each \( t \) and \( i \):
         \[
         \omega_{i,t} = \log \left( \sum_{a \in A \setminus 0} \exp(u_{i,a} + \beta E_{\omega_{t+1}}[\bar{V}_i(u_{i,a}, \omega_{t+1})|\omega_{i,t}]) \right)
         \]
      iii. Estimate AR(1) process (OLS): \( \omega_{i,t+1} = \rho_{i,0}^1 + \rho_{i,1}^1 \omega_{i,t} + \eta_{i,t+1} \)
      iv. Update value function:
         \[
         \bar{V}_i^1(u_0, \omega) = \log \left( \exp(u_0 + \beta E_{\omega'}[\bar{V}_i^0(u_0, \omega')|\omega]) \right) + \exp(\omega)
         \]
      v. Convergence check: \( ||\bar{V}^1 - \bar{V}^0|| < \epsilon_1, ||\rho_{i,0}^1 - \rho_{i,0}^0|| < \epsilon_2 \) and \( ||\rho_{i,1}^1 - \rho_{i,1}^0|| < \epsilon_3 \)
   (c) Calculate predicted market shares:
   \[
   \hat{s}_{jt} = s_{jt}(\delta^0, \theta) = \frac{1}{S} \sum_i \sum_{j_0} P_j(u_{i,j_0}, \omega_{it} | \alpha_i, \beta_i) w_{j_0}(\alpha_i, \beta_i)
   \]
   where \( P_j(u_{i,j_0}, \omega_{it} | \alpha_i, \beta_i) = \exp(v_0(u_{i,j_0}))/ (\exp(\omega_{it}) + \exp(v_0(u_{i,j_0}))) \)
   if \( j = j_0 \), and
   \[
   P_j(u_{i,j_0}, \omega_{it} | \alpha_i, \beta_i) = \frac{\exp(\omega_{it})}{\exp(\omega_{it}) + \exp(v_0(u_{i,j_0}))} \times \frac{\exp(v_{jt}(u_{jt}, \omega_{it}))}{\exp(\omega_{it})}
   \]
Applications: Video-camcorder market between 2000 and 2006

- **Main data-set:** Market shares and characteristics of 383 models and 11 brands, from March 2000 to May 2006

- **Auxiliary data:** Aggregate penetration and new-sales rates by years
Estimation results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Base Dynamic Model (1)</th>
<th>Dynamic Model without Repurchases (2)</th>
<th>Static Model (3)</th>
<th>Dynamic Model with Micro Moment (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean coefficients (α):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-.092 (.029)*</td>
<td>-.093 (7.24)</td>
<td>-6.86 (358)</td>
<td>-.367 (.065)*</td>
</tr>
<tr>
<td>Log price</td>
<td>-3.30 (1.03)*</td>
<td>-.543 (3.09)</td>
<td>-.099 (148)</td>
<td>-3.43 (.225)*</td>
</tr>
<tr>
<td>Log size</td>
<td>-.007 (.001)*</td>
<td>-.002 (.116)</td>
<td>-.159 (.051)*</td>
<td>-.021 (.003)*</td>
</tr>
<tr>
<td>(Log pixel)/10</td>
<td>.010 (.003)*</td>
<td>-.002 (.441)</td>
<td>-.329 (.053)*</td>
<td>.027 (.003)*</td>
</tr>
<tr>
<td>Log zoom</td>
<td>.005 (.002)*</td>
<td>.006 (.104)</td>
<td>.608 (.075)*</td>
<td>.018 (.004)*</td>
</tr>
<tr>
<td>Log LCD size</td>
<td>.008 (.002)*</td>
<td>.000 (.141)</td>
<td>-.073 (.098)</td>
<td>.004 (.005)</td>
</tr>
<tr>
<td>Media:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD</td>
<td>.035 (.006)*</td>
<td>.004 (1.16)</td>
<td>.074 (.332)</td>
<td>.060 (.019)*</td>
</tr>
<tr>
<td>Tape</td>
<td>.012 (.005)*</td>
<td>-.005 (.683)</td>
<td>-.667 (.318)*</td>
<td>.015 (.018)</td>
</tr>
<tr>
<td>Hard drive</td>
<td>.036 (.009)*</td>
<td>-.002 (1.55)</td>
<td>-.647 (.420)</td>
<td>.057 (.022)*</td>
</tr>
<tr>
<td>Lamp</td>
<td>.005 (.002)*</td>
<td>-.001 (.229)</td>
<td>-.219 (.061)*</td>
<td>.002 (.003)</td>
</tr>
<tr>
<td>Night shot</td>
<td>.003 (.001)*</td>
<td>.004 (.074)</td>
<td>.430 (.060)*</td>
<td>.015 (.004)*</td>
</tr>
<tr>
<td>Photo capable</td>
<td>-.007 (.002)*</td>
<td>-.002 (.143)</td>
<td>-.171 (.178)</td>
<td>-.010 (.006)</td>
</tr>
<tr>
<td>Standard deviation coefficients (Σ^{1/2}):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.079 (.021)*</td>
<td>.038 (1.06)</td>
<td>.001 (1,147)</td>
<td>.087 (.038)*</td>
</tr>
<tr>
<td>Log price</td>
<td>.345 (.115)*</td>
<td>.001 (1.94)</td>
<td>-.001 (427)</td>
<td>.820 (.084)*</td>
</tr>
</tbody>
</table>

Note.—Standard errors are in parentheses. All models include brand dummies, with Sony excluded. There are 4,436 observations.
* Statistically significant at the 5 percent level.

Implied elasticities:

• A market-wise \textit{temporary} price increase of 1\% leads to:
  
  – A contemporaneous decrease in sales of 2.55 percent
  
  – Most of this decrease is due to delayed purchases: 44 percent of the decrease in sales is recaptured over the following 12 months.

• The same market-wise \textit{permanent} price increase leads to a 1.23 percent decrease in demand (permanent).

• The difference is more modest when we consider a product-level price increase: 2.59 percent versus 2.41 percent (for the leading product in 2003).

• Why? Consumers substitute to competing brands when the change is permanent in about the same magnitude as the delayed response when the price change is temporary.
References


