Lecture Notes: Brand Loyalty and Demand for Experience Goods

Jean-François Houde
Cornell University

November 14, 2016

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

$$\begin{aligned} U_{ijt} &= X_{ijt}\beta + \lambda H_{ijt} + \epsilon_{ijt} \qquad (1) \\ \text{where} \qquad & \epsilon_{ijt} = \rho \epsilon_{ijt-1} + \nu_{ijt} \\ & H_{ijt} = \sum_{\tau=t_0}^{t-1} g(\tau) d_{ij\tau} \end{aligned}$$

- Empirical challenges:
  - Initial condition problem (i.e. d_{t_0} is endogenous and/or unobserved).
  - The presence of persistent unobserved heterogeneity can generates "spurious" state-dependence.
  - Multi-dimension integration when ϵ_{ijt} are correlated accross options and time:

$$\begin{aligned} \Pr(d_i | d_{t_0}) &= \Pr(U_{ijt} > U_{ikt}, \forall k \neq j, d_{ijt} = 1, t = 1...T) \\ &= \Pr(\epsilon_{ijt} - \epsilon_{ikt} > -(X_{ijt} - X_{ikt})\beta - (H_{ijt} - H_{ikt})\lambda, \\ &\qquad \forall k \neq j, d_{ijt} = 1, t = 1...T), \end{aligned}$$

    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:

$$U_{ij} = X_{ij}\beta + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \Omega) \tag{2}$$

Then the probability of choosing option 4 is:

$$\begin{aligned} \Pr(d_{i4} = 1) &= \Pr(\epsilon_{ij} - \epsilon_{i4} < -(X_{ij} - X_{i4})\beta, \forall k \neq 4) \\ &= \int_{A_1 \cap A_2 \cap A_3} dF(\epsilon_{i1} - \epsilon_{i4}, \epsilon_{i1} - \epsilon_{i3}, \epsilon_{i3} - \epsilon_{i4}) \\ &\quad \text{where } A_j = \{\epsilon_{ij} - \epsilon_{i4} < -(X_{ij} - X_{i4})\beta\} \end{aligned}$$

- Lets redefine the variables to compute the probability of choosing option 4:

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

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

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

$$\widehat{\Pr(d_{i4})} = \frac{\text{Number accepted draws}}{M} \tag{3}$$

  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:

$$\widehat{\Pr(d_{i4})} = \frac{1}{M}\sum_m \frac{1}{1+\sum_k \exp((-X^*_{ik}\beta - \nu^m_{ij})/\rho)} \tag{4}$$

...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 $\widehat{\Pr}(d_{i4}=1)$ we proceed sequentially:

  1. Draw ν^m_{i1}:
     Compute $\Phi_{i1} = \Pr(\nu_{i1} < -X^*_{i1}\beta)$. Draw η^m_1 from a truncated normal: $\eta \sim \Phi(-\frac{X^*_{i1}}{c_{11}})$
     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 ν^m_{i2} from a truncated normal (conditional on ν^m_{i1}):

$$\nu^m_2 = c_{12}\eta^m_1 + c_{22}\eta_2 < -X^*_{i2}\beta$$
$$\eta^m_2 \sim \Phi\Big(\frac{-X^*_{i2}\beta - c_{12}\eta^m_{i1}}{c_{22}}\Big) \equiv \Phi_{i2}$$

     Thus we first draw η^m_{i2} from $\Phi\Big(\frac{-X^*_{i2}\beta - c_{12}\eta^m_{i1}}{c_{22}}\Big)$ as before, and compute $\nu^m_{i2} = c_{12}\eta^m_{i1} + c_{22}\eta^m_{i2}$.

3. Compute Φ_{i3} similarly.
4. Compute $\widehat{\Pr}(d_{i4}=1)$:

$$\begin{aligned}\widehat{\Pr}(d_{i4}=1) \;=\; \frac{1}{M}\sum_{m}\Phi\Big(\frac{-X^{*}_{i1}\beta}{c_{11}}\Big)\times\Phi\Big(\frac{-X^{*}_{i2}\beta-c_{12}\eta^{m}_{i1}}{c_{22}}\Big)\\ \times\Phi\Big(\frac{-X^{*}_{i3}\beta-c_{12}\eta^{m}_{i1}-c_{23}\eta^{m}_{i2}}{c_{33}}\Big)\end{aligned}$$

- 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 ϵ_{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:

$$\begin{aligned}\widetilde{U}_{kt} &= U_{kt}-U_{j_t t}\\ \tilde{\epsilon}_{kt} &= \epsilon_{kt}-\epsilon_{j_t t}\\ \text{Where,}\quad & \tilde{\epsilon}\sim N(0,\widetilde{\Sigma}),\end{aligned}$$

  Where $\widetilde{\Sigma}=\widetilde{C}\widetilde{C}'$ is a $(J-1)T\times(J-1)T$ covariance matrix appropriately transformed to reflect the vector of choices $j_t, t=1...T$.
- With this transformation, $i\to j_{it}$ if $\widetilde{U}_{kt}\le 0,\forall k$ and $\tilde{\epsilon}_i=\widetilde{C}'\eta_i$.

- To compute $\widehat{\Pr}(j_i)$:

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

1: Draw η_{i11}^m from a truncated normal s.th:

$$\widetilde{U}_{i11}(\eta_{i11}^m) < 0.$$

2: Draw η_{i21}^m from a truncated normal s.th:

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

...

j_{i1}: Skip $\nu_{ij_{i1}1}^m$

...

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

1: Draw η_{i1t}^m from a truncated normal s.th:

$$\widetilde{U}_{i11}(\eta_{i11}^m, .., \eta_{iJt-1}^m, \eta_{i1t}^m) < 0.$$

...

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

$$\widetilde{U}_{ij_{i1}-11}(\eta_{i11}^m, .., \eta_{iJt-1}^m, \eta_{i1t}^m, ..., \eta_{ij_{it}-1t}^m < 0).$$

j_{i1}: Skip $\eta_{ij_{it}t}^m$

...

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

Back to Keane (1997): Modeling heterogeneity and state dependence in consumer choice behavior

- Keane (1997) estimates many different specifications of equation 1 using SMS with GHK and finds strong evidences of "true" state-dependence for Ketchup.
- The goal is to estimate a *statistical model* of product choice that flexibly account for state-dependence and *rich* unobserved preference heterogeneity
- Discrete-choice model:

$$\begin{aligned} U_{ijt} &= x_{it}\beta_j + p_{jt}(\phi_0 + x_{it}\phi_1 + \nu_i) + GL(H_{ijt},\alpha)\lambda + A^u_{jt}\Omega_{ijt} \\ \text{State dependence:} \quad & GL(H_{ijt},\alpha) = \begin{cases} \alpha GL(H_{ij,t-1},\alpha) + (1-\alpha)d_{ijt} & \text{If } t>1, \\ 0 & \text{If } t=1. \end{cases} \\ \text{Unobserved heterogeneity:} \quad & A^u\Omega_{ijt} = L_j W_i + \kappa\Gamma_{ij} + \delta_{ijt} \end{aligned}$$

- Decomposition of unobserved heterogeneity:
  - *Consumer heterogeneity:* $\kappa\Gamma_{ij}$ where $\Gamma_{ij} \sim N(0,I)$.
  - *Product taste heterogeneity:* $L_jW_i =$ Product FE $\times$ Unobserved taste for products, where $W_i \sim N(0,I)$.
  - *Time-varying preferences:* $\delta_{ijt} = \rho P_j \xi_{ij,t-1} + \rho\epsilon_{ij,t-1} + \eta_{ijt}$, where P_j is a vector of product FE, $\xi_{ijt} \sim N(0,1)$, $\eta_{ijt} \sim N(0,1)$ and $\epsilon_{ijt} \sim N(0,1)$.
- **Estimation:** Method of Simulated Moments (MSM)

$$m(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} W_{ijt}\left[d_{ijt} - \hat{P}_{GHK}(d_{ijt}|d_{ij1},\ldots,d_{ij,t-1},X_i,A,P,\theta)\right] = 0$$

$$\text{Where,} \quad \hat{P}_{GHK}(d_{ijt}|d_{ij1},\ldots,d_{ij,t-1},X_i,A,P,\theta) = \frac{\hat{P}_{GHK}(d_{ij1},\ldots,d_{ij,t})}{\hat{P}_{GHK}(d_{ij1},\ldots,d_{ij,t-1})}$$

Estimation Results: Basic Specifications

Table 2. The Interaction of Heterogeneity and State Dependence in Some Simple Models With Independent Time-Varying Error Components

Parameter name	*Model 1: GL form of state dependence but no heterogeneity*	*Model 2: Kappa heterogeneity only and no state dependence (similar to Dirichlet multinomial)*	*Model 3: Kappa-type heterogeneity and GL form of state dependence*	*Model 4: Kappa heterogeneity, a common factor, and GL form of state dependence*
Kappa		.915 (.058)	.444 (.052)	.221 (.040)
Lambda	1.985 (.106)		.889 (.154)	1.096 (.160)
Alpha	.813 (.015)		.833 (.029)	.836 (.026)
L1				−.361 (.084)
L2				−.531 (.102)
L3				.484 (.086)
L4				1.148 (.143)
L5				.897 (.115)
L6				.170 (.072)
Hunt's	−.428 (.023)	−.559 (.030)	−.526 (.028)	−.584 (.030)
Del Monte	−.644 (.030)	−.821 (.038)	−.763 (.036)	−.870 (.050)
Plastic	.330 (.023)	.347 (.030)	.373 (.028)	.391 (.029)
Ounces	−.024 (.005)	−.016 (.005)	−.023 (.005)	.001 (.008)
Oz.**2/100	.007 (.005)	−.009 (.006)	.000 (.006)	−.035 (.010)
Price	−1.453 (.053)	−1.663 (.057)	−1.656 (.058)	−1.695 (.067)
Line ad	.485 (.105)	.397 (.094)	.425 (.101)	.439 (.106)
Display ad	.718 (.043)	.583 (.038)	.638 (.042)	.698 (.044)
Major ad	.803 (.049)	.681 (.042)	.759 (.047)	.788 (.052)

- **Model 2:** Nearly 50% of the *error* is due to time-invariant consumer heterogeneity (κ)
- **Model 3-4:** Adding state-dependence reduces the importance of unobserved heterogeneity to 30%, and heterogeneity in taste across brand (L_j) further reduces it to less than 20% (between $11 - 18\%$).

Estimation Results: Richest Specifications

Table 4. Models With Heterogeneity/Serial Correlation and State Dependence

Parameter name	*Model 11: Let the ε_{ijt} be correlated across brands (P)*	*Model 12: Add correlation in random effects across brands (L)*	*Model 13: Let the ε_{ijt} be correlated over time, with AR(1) parameter rho*	*Model 14: Add observed heterogeneity in intercepts*	*Model 15: Add observed heterogeneity in price response*	*Model 16: Add unobserved heterogeneity in price response*
Kappa	.646 (.106)	.529 (.089)	.431 (.081)	.475 (.091)	.501 (.091)	.549 (.092)
Lambda	1.456 (.193)	1.120 (.214)	1.355 (.336)	1.433 (.426)	1.391 (.356)	1.346 (.379)
Alpha	.850 (.021)	.868 (.030)	.897 (.034)	.911 (.033)	.907 (.033)	.909 (.037)
L1		−.342 (.110)	−.246 (.106)	−.255 (.106)	−.248 (.113)	.090 (.103)
L2		−.854 (.147)	−.406 (.129)	−.139 (.144)	−.028 (.155)	.303 (.151)
L3		.707 (.123)	1.180 (.180)	1.490 (.190)	1.664 (.211)	1.857 (.237)
L4		1.876 (.317)	2.007 (.323)	2.550 (.378)	2.599 (.368)	2.454 (.320)
L5		1.385 (.213)	1.129 (.167)	.669 (.167)	.699 (.178)	−1.959 (.328)
L6		.434 (.134)	.637 (.141)	.826 (.144)	.976 (.158)	1.285 (.192)
P1	1.419 (.224)	.544 (.179)	.361 (.196)	.317 (.233)	.434 (.210)	.276 (.168)
P2	2.041 (.314)	.540 (.228)	.976 (.229)	1.621 (.265)	1.755 (.259)	1.280 (.198)
P3	−1.224 (.128)	−.276 (.154)	1.871 (.283)	2.211 (.270)	2.456 (.292)	2.796 (.307)
P4	1.663 (.302)	2.047 (.371)	2.092 (.444)	2.670 (.407)	2.776 (.407)	2.543 (.385)
P5	1.124 (.252)	1.183 (.340)	−.241 (.311)	−.731 (.380)	−.653 (.387)	−.902 (.361)
P6	3.345 (.323)	1.841 (.203)	1.946 (.236)	2.118 (.344)	2.341 (.253)	2.804 (.311)
Rho			.139 (.032)	.171 (.031)	.167 (.031)	.134 (.033)
Hunt's	−.202 (.045)	−.492 (.049)	−.515 (.060)	−.584 (.105)	−.470 (.110)	−.465 (.099)
Del Monte	−.581 (.069)	−.916 (.076)	−.778 (.069)	−.848 (.134)	−.715 (.141)	−.692 (.121)
Plastic	.147 (.042)	.322 (.042)	.454 (.048)	.618 (.104)	.569 (.119)	.572 (.139)
Ounces	−.102 (.009)	.005 (.018)	.004 (.023)	−.058 (.026)	−.061 (.028)	−.042 (.032)
Oz.**2/100	.015 (.011)	−.050 (.026)	−.046 (.029)	−.009 (.036)	−.001 (.037)	−.015 (.038)
Price	−1.943 (.118)	−1.895 (.097)	−2.052 (.105)	−2.056 (.111)	−1.898 (.241)	−2.355 (.261)
Price*inc					.011 (.003)	.023 (.005)
Price*HH size					−.143 (.062)	−.206 (.072)
Var(**P**)						1.818 (.314)

- **Model 12:** Adding product-specific time-varying heterogeneity suggest rich dynamic switching/loyalty patterns. The contribution time-invariant consumer heterogeneity is roughly 10% of the error.
- **Model 13-16:** The AR(1) coefficient is significantly different from zero (reject *constant* correlation in taste over time).
- **Model 13-16:** The state-dependence parameters ($\lambda = 1.3$ and $\alpha = 0.9$) are fairly stable across specifications, and suggest that brand loyalty is an important phenomenon. A one-time switch is equivalent to a short-run price increase of roughly 5 cents.
- **Model 16:** Accounting for rich heterogeneity is crucial to accurately measure brand loyalty. A one-time switch leads to the following drop in utility:

	t	$t+1$	$t+2$		$t+10$
Model 1	0.37	0.3	0.24	...	0.047
Model 16	0.11	0.101	0.092	...	0.047

- **Model 16:** Allowing for state-dependence and serial-correlation substantially reduce the importance consumer heterogeneity (κ): between 4% and 22% of the residual variance.

Ackerberg (2001): Informative versus persuasive advertising

- **Question:** How to distinguish between informative and persuasive advertisement?
- **Two views:**
  - *Information:* Ads inform consumers about the quality and/or price of products
  - *Persuasion:* Ads interact with consumption of brands (i.e. complementarity), and create a prestige effect that increases demand for the good.
- **Why do we care?**
  - The two theories have different implications for the welfare effect of advertising (wasteful investment?)
  - From an antitrust point of view, the two theories have different implications for the effect of advertising on market-power.
  - Informative advertising typically has a *pro-competitive* effect, while persuasive advertising can create barriers to entry.
- **Identification problems:**
  - Even with the **best** natural experiment, the effect of ads on demand does not tell us much about the underlying **mechanism**.
  - **Solution:** Exploit fairly long panel of consumers' repeated purchases to measure the effect of advertising on *experienced* and *inexperienced* consumers.
  - **Initial conditions?**
    - Usually, the experience level of consumers is highly endogenous.
    - Ackerberg solves this problem by studying the life-cycle of a new product (Yoplait 150) from its introduction to its steady-state.

- **Simple test:** Effect of ads on aggregate *new* purchases versus *repeated* purchased.

	Dependent Variable: Initial Purchases				Dependent Variable: Repeat Purchases			
	1	2	3	4	1	2	3	4
N	918	918	678	918	918	918	678	918
R^2	.066	.085	.107	.066	.162	.149	.120	.162
Market	.222	.002	.224	.223	.700	.006	.832	.700
Dummy	(.062)	(.000)	(.069)	(.062)	(.089)	(.000)	(.111)	(.089)
Price	−5.298	−.038	−7.388	−5.354	−3.954	−.029	−5.512	−3.942
	(1.568)	(.013)	(1.726)	(1.585)	(1.829)	(.014)	(2.207)	(1.838)
Ads	.044	.030	.042	.044	.020	.014	.014	.016
	(.022)	(.015)	(.021)	(.022)	(.023)	(.017)	(.024)	(.024)
t-value	1.981	1.925	2.046	1.988	.873	.818	.596	.679

- **Discrete-choice model:** Repeated logit with random effect

$$U_{ijt} = \begin{cases} \alpha_i + x_i\beta_x - \beta_p p_{jt} + \epsilon_{ijt} & \text{If } j = 1 \\ \epsilon_{i0t} & \text{If } j = 0 \end{cases}$$

- Likelihood function:

$$L(d_i, x_i|\theta) = \int \prod_{t=1}^{T} \Pr(d_{it}|x_{it}, \alpha_i, \theta) f(\alpha_i)\alpha_i$$

- Results:

Parameter	Simple Logit	Normal Random Effect	Simple Logit	Normal Random Effect	Flexible Ad Coefs	.5 Logit	With Mean Advertising	Extra Promotional Variables
Advertising ∗	2.04073	2.30566	—	—	2.32360	—	—	—
Inexperienced	(.72313)	(.77561)			(.78683)			
Advertising ∗	.90371	.43304	—	—	1.33200	—	—	—
Experienced	(.63504)	(1.21180)			(1.39850)			
t-statistic on diffference	1.47662	1.58703	—	—	—	—	—	—
Advertising	—	—	1.71550	2.01370	—	2.10570	1.73080	2.40619
			(.76392)	(.79037)		(.85627)	(.82047)	(.89738)
Advertising ∗	—	—	−.14812	−.35627	−.29487	−.27106	−.35253	−.39207
Num prev pur			(.06282)	(.10803)	(.12079)	(.14411)	(.10904)	(.11248)
Mean	—	—	—	—	—	—	2.48400	—
ads							(2.40050)	

Erdem and Keane (1996): Decision-Making under uncertainty

- Estimate a dynamic bayesian learning model of demand for detergent using the A.C. Nielsen **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:

  `https://research.chicagobooth.edu/kilts/marketing-databases/nielsen`

- Why?
  - Market with frequent brand introduction.
  - Reasonable to assume that consumers learn about the quality of the product **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:

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

  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. Atribute of good j if "experienced" at t:

$$A_{jt}^E = A_j + \delta_{jt}.$$

  2. Utility after realization of δ_{jt} and e_{jt}:

$$\begin{aligned} U_{jt} &= -\omega_p P_{jt} + \omega_A A_{jt}^E - \omega_A r A_{jt}^{E^2} + e_{jt} \\ \text{Where} \quad & r = \text{risk aversion coefficient} \\ & e_{jt} = \text{"logit" utility shock} \\ & P_{jt} = \text{Price of } j \text{ (stochastic)} \end{aligned}$$

  3. Expected utility:

$$\begin{aligned} E[U_{jt}|I(t)] &= -\omega_p P_{jt} + \omega_A E[A_{jt}^E|I(t)] - \omega_A r E[A_{jt}^{E^2}|I(t)] - \\ & \qquad \omega_A r E\Big[(A_{jt}^E - E[A_{jt}^E|I(t)])^2]|I(t)\Big] + e_{jt} \end{aligned}$$

- Outside options:

$$\begin{aligned} E[U_{0t}|I(t)] &= \Phi_0 + \Psi_0 t + e_{0t} \\ E[U_{NPt}|I(t)] &= \Phi_{NP} + \Psi_{NP} t + e_{NPt} \end{aligned}$$

- 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: δ is mean zero.

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

  3. Advertising message (with probablity 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 σ_η^2

  * Advertising is exogenous and strictly informative: η_{jt} is mean zero.

- Bayesian Updating: Update expectation about good j's attribute
$$\begin{aligned} E[A_j|I(t)] \;=\; & E[A_j|I(t-1)] + d_{jt}\beta_{1jt}\big[A_{jt}^E - E[A_{jt}^E|I(t-1)]\big] \\ & + a d_{jt}\beta_{2jt}\big[S_{jt} - E[S_{jt}|I(t-1)]\big] \end{aligned}$$
  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:
$$\begin{aligned} \nu_j(t) \;=\; & \nu_j(t-1) + d_{jt}\beta_{1jt}\big[-\nu_j(t-1) + \delta_{jt}\big] \\ & + a d_{jt}\beta_{2jt}\big[-\nu_j(t-1) + \eta_{jt}\big] \end{aligned} \tag{5}$$
  With $\nu_j(0) = A - A_j, \forall j$.

Similarly, the precision of signals is updated using the following markov process:

$$\sigma_{\nu_j}(t) = \left[\frac{1}{\sigma_\nu(0)} + \frac{\sum_{s \leq t} d_{js}}{\sigma_\delta^2} + \frac{\sum_{s \leq t} ad_{js}}{\sigma_\nu^2} \right]^{-1} \tag{6}$$

- Timing:

$t = 0$ $*$ Purchasing decision based on $\nu_j(0)$.
$*$ New signals: δ_{j0} (if $d_{j0} = 1$) and η_{j0} (if $ad_{j0} = 1$).
$*$ Update $\nu_j(1)$ according to equation 5

$t = 1$ $*$ Purchasing decision based on $\nu_j(1)$.
$*$ New signals: δ_{j1} (if $d_{j1} = 1$) and η_{jt} (if $ad_{j1} = 1$).
$*$ Update $\nu_j(2)$ according to equation 5

...

- Therefore at any period t the payoff relevant state variables are:

$$I(t) = \Big\{ \underbrace{\sum_{s \leq t-1} d_{js}, \sum_{s \leq t-1} ad_{js}}_{Discrete}, \underbrace{\nu_{jt}}_{Continuous} \Big\}_{j=1...J}$$

- 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:

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

  - Integration is complicated by the fact that ν_{jt} is serially correlated: Must integrate the sequence of past ν_{js}'s using simulation method: simulate M sequences of ν_{jt}, just like in Hendel and Nevo (2006).

– 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 $I(0) = \{0, 0, \nu_{j0}\}_{j=1...,J}$, 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 $\mathcal{I}^*(t)$ of the states and interpolate between them using least-squares to compute $EV(I(t))$ at $I(t) \notin \mathcal{I}^*$.
- Backward induction algorithm for a fix grid $\mathcal{I}^*$:

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

$$\begin{aligned} EV_T(I(T)) &= \int \left\{ \max_j EU_{jT}(I(T)) + e_{jT} \right\} dF(e) \\ &= \log \left(\sum_j \exp \left(EU_{jT}(I(T)) \right) \right) \end{aligned}$$

2. Run the following regression:

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

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, polynomials, etc).

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

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

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

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

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

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

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 advertizing 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.
- Forward looking model fits better...

Table 3 Structural Model Estimates

	Immediate Utility Maximization[1] ($\gamma = 0$)		Forward-looking Dynamic Structural Model[2] ($\gamma = 0.995$)	
Parameter	Estimate	t-statistic	Estimate	t-statistic
price coefficient ($-w_p$)	−0.790	−12.26	−0.795	−12.31
utility weight (w_A)	28.356	1.73	34.785	1.84
risk coefficient (r)	3.625	2.08	4.171	2.25
initial variance ($\sigma_\nu^2(t)$)	0.053	4.64	0.040	4.21
mean attribute levels (A_j):				
A_{Dash}	0.049	0.74	0.040	0.74
A_{Cheer}	0.019	0.27	0.012	0.21
A_{Solo}	0.056	0.84	0.047	0.87
A_{Surf}	0.105	1.65	0.089	1.77
A_{Era}	0.137	2.41	0.120	2.64
A_{Wisk}	0.040	0.59	0.029	0.53
A_{Tide}	0.138	–	0.120	–
"Other Brands" intercept (Φ_O)	−17.657	−7.98	−17.267	−7.59
"Other Brands" time trend (Ψ_O)	0.018	8.53	0.018	8.91
"No Purchase" intercept (Φ_{NP})	−15.408	−6.99	−19.537	−8.55
"No Purchase" time trend (Ψ_{NP})	0.011	3.17	0.012	3.42
experience variability (σ_δ)	0.374	9.17	0.33	8.37
advertising variability (δ_ξ)	3.418	6.29	3.08	5.57

[1] −LL = 7312.09 AIC = 7324.09 BIC = 7384.49
[2] −LL = 7306.05 AIC = 7322.05 BIC = 7378.45

Crawford and Shum (2005): Uncertainty and Learning in Pharmaceutical Demand

TABLE II

SWITCHING PROBABILITIES OVER THE COURSE OF TREATMENT[a]

Prescription Number	Total Treatment Length					
	5	6	7	8	9	10
2	14.3	13.6	10.9	10.0	7.8	9.2
3	11.6	11.6	6.3	8.8	7.8	6.6
4	8.9	5.6	5.4	3.1	7.8	3.9
5	13.4	7.9	10.0	8.8	4.9	5.3
6		11.3	6.3	5.7	2.9	5.3
7			9.5	10.0	7.8	11.8
8				8.1	4.9	11.8
9					7.8	5.3
10						11.8

[a]The (i, j)th entry is the percentage of treatment sequences of length j in which a switch was observed during the ith $(i \leq j)$ prescription.

- **Quote:** 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:

$$\begin{aligned} x_{jnt} &\sim N(\mu_{jn}, \sigma_n^2) \\ y_{jnt} &\sim N(\nu_{jn}, \tau_n^2) \end{aligned}$$

- Initial priors about the match values:

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

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

- Expected Utility (CARA):

$$\begin{aligned} u(x_{jnt}, p_n, \epsilon_{jnt}) &= -\exp(-r x_{jnt}) - \alpha p_n + \epsilon_{jnt} \\ \widetilde{EU}(\mu_{jn}(t), \nu_{jn}(t), p_n, \epsilon_{jnt}) &= -\exp\Big(-r\mu_{jn}(t) + 1/2 r^2(\sigma_n^2 + V_{jn}(t))\Big) \\ &\quad -\alpha p_n + \epsilon_{jnt} \\ &= EU(\mu_{jn}(t), V_{jn}(t), p_n) + \epsilon_{jnt} \end{aligned}$$

- 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:

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

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

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

$$\begin{aligned} V(s) &= \int \max_n EU(s) + \epsilon_n + \beta E\big[(1-h(s'))V(s')|d_n = 1, s\big] \\ &= \log\Big[\sum_n \exp\Big(EU(s) + \beta E\big[(1-h(s'))V(s')|d_n = 1, s\big]\Big)\Big] \end{aligned}$$

- **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 dug n for each $s \in S^*$:
$$E\Big[V(s|d_n = 1, s\Big] = \frac{1}{M}\sum_m (1-h(s^m))V^0(s^m)$$
     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 $s \in S^*$:
$$V^1(s) = \log\Big[\sum_n \exp\Big(EU(s) + \beta E\Big[V(s|d_n = 1, s\Big]\Big)\Big]$$
  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).

Figure 1: Results

TABLE III

DYNAMIC MODEL: PARAMETER ESTIMATES

Parameter	Est.	Std. Err.	Est.	Std. Err.
Illness heterogeneity distribution	Recovery Probability		Type Probability	
θ_1 (Type 1)	0.433	0.003	0.593	0.006
θ_2 (Type 2)	0.127	0.003	0.335	0.006
θ_3 (Type 3)	0.199	0.007	0.043	0.001
θ_4 (Type 4)	0.432	0.011	0.029	0.002
Means, symptom match values[b]	Type 1		Type 2	
$\underline{\mu}_1$	0.927	0.282	1.195	0.369
$\underline{\mu}_2$[c]	0.928	0.287	0.428	0.166
$\underline{\mu}_3$	0.481	0.197	−0.028	0.178
$\underline{\mu}_4$	0.335	0.161	−0.145	0.079
$\underline{\mu}_5$	0.451	0.174	−0.483	0.137
Means, curative match values[b]	Type 1		Type 2	
$\underline{\nu}_1$	0.014	0.003	0.006	0.000
$\underline{\nu}_2$[c]	0.015	0.005	0.006	0.001
$\underline{\nu}_3$	0.013	0.030	0.006	0.095
$\underline{\nu}_4$	0.013	0.084	0.014	0.009
$\underline{\nu}_5$	−0.034	0.000	−0.038	0.000
Std. dev., symptom match values				
$\underline{\sigma}$	1.574	0.448		
Std. devs., symptom signals				
σ_1	0.998	0.287		
σ_2	1.134	0.326		
σ_3	1.375	0.395		
σ_4	1.159	0.333		
σ_5	0.931	0.268		
Std. dev., curative match values				
$\underline{\tau}$	0.007	0.000		
Std. dev., curative signals				
τ	0.007	0.001		
Price coefficient, α[a]	1.080	0.091		
Risk-aversion parameter, r	0.990	0.274		
Discount rate, β	0.950	Fixed		
Number of observations	34,972			
Number of similar draws	30			
Log likelihood function	−124,484.34			

[a]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.

References

Ackerberg, D. (2003). Advertising, learning, and consumer choice in experience good markets: A structural empirical examination. *International Economic Review 44*, 1007–1040.

Ackerberg, D. A. (2001, Summer). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics 32*(2), 316–333.

Crawford, G. and M. Shum (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica 73*, 1137–1174.

Erdem, T. and M. P. Keane (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science 15*(1), 1–20.

Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer stockpiling behavior. *Econometrica 74*(6), 1637–1673.

Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Review of Economics and Statistics 15*(3), 310–327.

Keane, M. P. and K. I. Wolpin (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics 76*(4), 648–672.