Recent Developments and Challenges of Single Agent Dynamic Models

Andrew Ching
Rotman School of Management
University of Toronto

First version: June 3, 2009
This extended version: Nov 23, 2009

Roadmap

- Dynamic Structural Models in Marketing: Reward Programs, Learning, Stockpiling, Technology Adoption.
- Standard estimation approach - Nested Fixed Point Algorithm (Rust, 1987)
- Bayesian estimation approach - Modified MCMC algorithm (Imai, Jain and Ching, 2009)
- Simplified approach - Erdem, Imai and Keane (2009) and Hendel and Nevo (2006)
- Another approximation approach for estimating models with product level data - Ching (2000, 2008)
- Measuring Time Preference and Time Discounting.
- Conclude.

Rewards Program

- Lewis (2006), Hartmann and Viard (2008) are the first who use dynamic programming models to study the incentive of loyalty programs.
- A consumer needs to earn enough points to redeem a reward (e.g., free game, gifts, or air-tickets)
- A nice environment to measure consumer forward-looking incentives.
- Challenges: Some rewards programs, like Air Miles, could be quite complicated to study.

Model (An Example)

A simple dynamic model of store choice with reward programs.

- There are two supermarkets in a city $(j = 1, 2)$.
- Each store offers a stamp card, which can be exchanged for a gift upon completion.
- Consumers get one stamp for each visit with a purchase.
- Reward programs at the two supermarkets differ in terms of: (i) the number of stamps required for a reward $(\bar{S}_j)$, and (ii) the mean value of the reward (G_j).
- $G_{ij} \sim N(G_j, \sigma^2_{G_j})$, i indexes consumers.
- Once consumers receive a gift, they will start with a blank stamp card again in the next period.

Model (cont'd)

- $p_{ijt} \sim i.i.d.\ N(\underline{p}, \sigma_p^2)$.
- Let $s_i = (s_{i1}, s_{i2})$ be the number of stamps collected.
- Consumer i's single period utility of visiting supermarket j is:

$$U_{ijt}(s_{it}, p_{it}; \theta) = \begin{cases} \alpha_j - \gamma p_{ijt} + G_{ij} + \epsilon_{ijt} & \text{if } s_{ijt} = \bar{S}_j - 1, \\ \alpha_j - \gamma p_{ijt} + \epsilon_{ijt} & \text{otherwise.} \end{cases}$$

- The utility of "not shopping" is normalized to zero, i.e., $U_{i0t}(s) = 0$.
- We assume that ϵ follows i.i.d. extreme value distribution.

Model (cont'd)

- The consumer i's objective is to choose a sequence of store choices to maximize the sum of the present discounted future utility.

$$\max_{\{d_{ijt}\}_{t=1}^{\infty}} E\left[\sum_{t=1}^{\infty} \beta^{t-1} d_{ijt} U_{ijt}(s_{it}, p_{it}; \theta)\right]$$

where $d_{ijt} = 1$ if consumer i chooses store j in period t and $d_{ijt} = 0$ otherwise, and β is the discount factor.

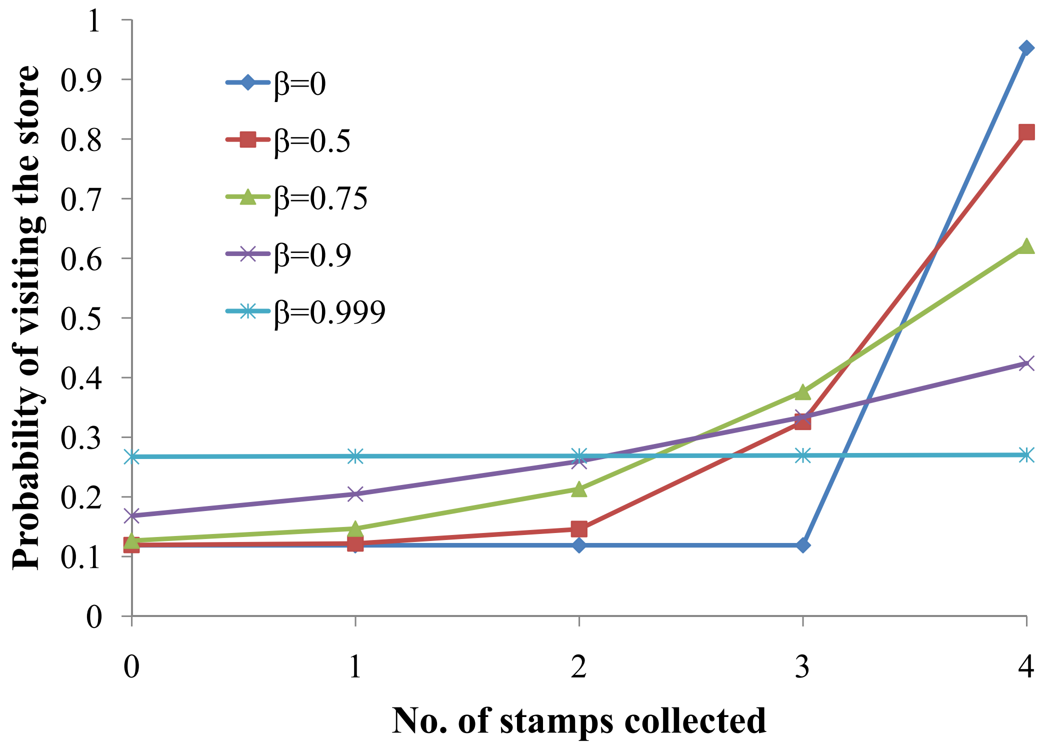
- The evolution of the state, s_{it}, is as follows.

$$s_{ijt+1} = \begin{cases} s_{ijt} + 1 & \text{if } d_{ijt} = 1 \text{ and } s_{ijt} < \bar{S}_j - 1 \\ 0 & \text{if } d_{ijt} = 1 \text{ and } s_{ijt} = \bar{S}_j - 1 \\ s_{ijt} & \text{if } d_{ijt} = 0. \end{cases}$$

Implications of the Model

- The main dynamics of the model is the intertemporal trade-off created by the reward program.
- Suppose that a consumer is close to completion of the stamp card for store 1, and store 2 offers a lower price today.
- If the consumer chooses store 2 today, he/she will get a better deal today but will delay the completion of the stamp card for store 1.
- If β is less than one, the delay will lower the present discounted value of the gift.
- The incentive not to delay the completion will increase as he/she gets closer to the completion of the stamp card.

Choice Probabilities



- Only one store and the outside option.
- Other parameters are set as $\bar{S}_1 = 5$, $\alpha_1 = -2.0$, $G_1 = 3.0$, $\sigma_{G_1} = 0$, and $\gamma = -1.0$.

Bellman Equation

$$
\begin{aligned}
&V(s_i; p_i, G_i, \theta) \\
&\quad \equiv \quad E_\epsilon \max\{V_1(s_i; p_{i1}, G_i, \theta) + \epsilon_{i1}, V_2(s_i; p_{i2}, G_i, \theta) + \epsilon_{i2}, V_0(s_i; G_i, \theta) + \epsilon_{i0}\} \\
&\quad = \quad \log(\exp(V_1(s_i; p_{i1}, G_i, \theta)) + \exp(V_2(s_i; p_{i2}, G_i, \theta)) + \exp(V_0(s_i; G_i, \theta))).
\end{aligned}
$$

The second equality follows from the extreme value assumption on ϵ.

$$
\begin{aligned}
V_j(s_i; p_{ij}, G_i, \theta) &= \begin{cases} \alpha_j - \gamma p_{ij} + G_{ij} + \beta E_{p'} V(s'; p', G_i, \theta) & \text{if } s_{ij} = \bar{S}_j - 1, \\ \alpha_j - \gamma p_{ij} + \beta E_{p'} V(s'; p', G_i, \theta) & \text{otherwise,} \end{cases} \\
V_0(s_i; G_i, \theta) &= \beta E_{p'} V(s'; p', G_i, \theta)
\end{aligned}
$$

The expectation with respect to p' is defined by

$$
E_{p'} V(s'; p', G_i, \theta) = \int V(s'; p', G_i; \theta) dF(p').
$$

Numerical Solution of the Model

- We use contraction mapping to solve for V. Start with an initial guess, V^0. Then use Bellman operator to update V^n recursively until it converges (Refer this to standard DP step).

$$\hat{E}_p[V^{n-1}(s';p',G_i,\theta)] = \frac{1}{M}\sum_{m=1}^{M} V^{n-1}(s';p^m,G_i,\theta)$$

$$V^n(s_i;p_i,G_i,\theta) = E\max_j\{U_{ij}(s_i;p_i,\theta) + \beta\hat{E}_{p'}[V^{n-1}(s';p',G_i,\theta)]\},$$

where $\{p^m\}_{m=1}^{M}$ are drawn from the price distribution.

Nested Fixed Point Algorithm (Rust, 1987)

- Likelihood (i.e., choice probabilities in this case) for a given θ is a function of the alternative specific value functions evaluated at θ.
- In maximum likelihood, the parameter estimate, $\hat{\theta}$, maximize the likelihood of the observed data.
- If we use derivative based maximization routine to search for $\hat{\theta}$, and suppose that there are k parameters, we need to evaluate the likelihood at $k+1$ different θ (because we calculate the numerical partial derivatives). In other words, we need to numerically solve the value functions $k+1$ times in each iteration. If it takes 200 iterations to converge and there are 20 parameters to estimate, we need to solve the value function numerically 40,000 times.
- In bayesian MCMC estimation, the number of iterations required for convergence is even much higher, typically, more than 30,000 iterations.

Highlights of IJC method

- In the nested fixed point algorithm (Rust 1987), the value functions need to be solved at each trial parameter vector.
- IJC make use of the past outcomes of the estimation algorithm, $\{V^l(s, \epsilon^l; \theta^l)\}_{l=l-N}^{r-1}$, to approximate $E_\epsilon[V(s, \epsilon; \theta)]$ (take advantage of the continuity of V in θ). In contrast, the conventional method simply throws away this information.
- Applicable to general infinite horizon stationary DDP models.
- It estimates and solves the DDP model simultaneously.
  - Computational burden of IJC algorithm of each iteration is comparable to that of estimating a standard static discrete choice model.
- It can incorporate unobserved heterogeneities (random coefficients).

Key features of IJC method

- MCMC algorithm that bases on a proxy of the value function.
- Consider a model without unobserved consumer heterogeneity.
- The outputs of the algorithm in iteration r include $\{\theta^r, E_p\tilde{V}^r(s,p;\theta^r)\}$. Store $H^r = \{\theta^l, E_pV^l(s,p;\theta^l)\}_{l=r-N}^{r-1}$.
- Use Metropolis-Hastings within Gibbs to obtain θ^r.
- Given a candidate draw θ^{*r}, use $\{E_p\tilde{V}^l(s,p;\theta^l)\}_{l=r-N}^{r-1}$ to obtain an estimate of $\hat{E}_{p'}[V(s',p';\theta^{*r})]$.
- Make M draws of prices, $p^m = (p_1^m, p_2^m)$, from the price distribution. Obtain $\tilde{V}^r(s,p^m;\theta^{*r})$ by applying the Bellman operator only once based on $\hat{E}_{p'}[V(s',p';\theta^{r*})]$. Then,

$$E_p\tilde{V}^r(s,p;\theta^{*r}) = \frac{1}{M}\sum_{m=1}^{M}\tilde{V}^r(s,p^m;\theta^{*r}).$$

IJC's DP step

IJC propose to replace the standard DP step with the following step:

- Let $H^r = \{\theta^{*l}, E_p\tilde{V}^l(., p; \theta^{*l})\}_{l=r-N}^{r-1}$ be the history of simulated draws up to the previous iteration $r-1$.
- For each s,

$$\hat{E}_{p'}[V(s', p'; \theta^{*r})] = \sum_{n=1}^{N} E_p\tilde{V}^{r-n}(s', p, \theta^{*r-n})\frac{K_h(\theta^{*r} - \theta^{*r-n})}{\sum_{k=1}^{N} K_h(\theta^{*r} - \theta^{*r-k})},$$

where $K_h()$ is a kernel with bandwidth $h > 0$.

- Simulate $\{p^m\}_{m=1}^M$, and compute the pseudo-value function

$$\tilde{V}^r(s, p^m; \theta^{*r}) = \max_j\{U_j(s, p^m; \theta^{*r}) + \beta\hat{E}_{p'}[V(s', p'; \theta^{*r})]\},$$

$$E_p\tilde{V}^r(s, p; \theta^{*r}) = \frac{1}{M}\sum_{m=1}^{M}\tilde{V}^r(s, p^m; \theta^{*r}).$$

The Pseudo-Value Function

- The pseudo-value function is defined as follows.

$$\begin{aligned}
&\tilde{V}^r(s, p^m; \theta^{*r}) \\
&\quad \equiv \quad E_\epsilon \max\{\tilde{V}_0(s; \theta^{*r}) + \epsilon_0, \tilde{V}_1(s; p_1^m, \theta^{*r}) + \epsilon_1, \tilde{V}_2(s; p_2^m, \theta^{*r}) + \epsilon_2\} \\
&\quad = \quad \log(\exp(\tilde{V}_0^r(s; \theta^{*r})) + \exp(\tilde{V}_1^r(s, p_1^m; \theta^{*r})) + \exp(\tilde{V}_2^r(s, p_2^m; \theta^{*r})))
\end{aligned}$$

where

$$\begin{aligned}
\tilde{V}_j^r(s, p_j^m; \theta^r) &= \begin{cases} \alpha_j - \gamma p_j^m + G_j + \beta \hat{E}_{p'}^r V(s', p'; \theta^r) & \text{if } s_j = \bar{S}_j - 1, \\ \alpha_j - \gamma p_j^m + \beta \hat{E}_{p'}^r V(s', p'; \theta^r) & \text{otherwise,} \end{cases} \\
\tilde{V}_0^r(s; \theta^r) &= \beta \hat{E}_{p'}^r V(s', p'; \theta^r).
\end{aligned}$$

Remark 1

When S is discrete, and N increases at an appropriate rate as l increases, IJC show that:

- Theorem 1: The sequence of approximated value function V^l converges to V in probability uniformly over (S, ϵ, Θ).
- Theorem 2: θ^l converges to the true posterior distribution in total variation norm.
- Note that in practice, we need to fix N.

Remark 2

- The approximation step of the Emax functions can also be applied in the classical methods (Brown and Flinn 2006).
- However, applying it in the Bayesian estimation has at least two advantages:
  - Past value functions are evaluated at parameter vectors, which are randomly distributed around the current parameter value – good for non-parametric approximation. This is typically not the case in the classical methods.
  - Accurately simulating a posterior distribution is usually easier than finding the global maximum/minimum of a complex likelihood/GMM objective function.

How to choose N and h

- The rate of convergence is usually faster for smaller N, because the set of past pseudo-value functions used for approximation are more recent and accurate.
- However, smaller N also leads to the weighted average of the pseudo-value functions more sensitive to the change in the set. As a result, we may get larger posterior standard deviation.
- One trick we can use is to set N to be small at the beginning and increase it as the number of iterations increases.

How to choose N and h (cont'd)

- A general strategy to choose N and h (bandwidth): check how good the approximation for the value function is during the estimation.
  - For instance, in every 1,000 iterations, compute the means of the MCMC draws, $\bar{\tilde{\theta}}$, and compare the distance between the pseudo-value function and the exact value function at $\bar{\tilde{\theta}}$.
  - If memory is not a constraint, store a large N past pseudo-value functions and use the most recent $N' < N$ to do the Emax approximation.
  - One can increase N' as soon as the researcher discovers that the approximation is not good enough.
  - Then set a new h based on standard optimal bandwidth formula, e.g., Silverman's rule of thumb.

Estimation 1: Impact of N

- We simulate samples based on the solution of the dynamic model. We then estimate the parameters of the model by using different N's.
- True parameters of the simulation
  - Store intercepts: $\alpha_1 = \alpha_2 = 0$.
  - Population mean of rewards: $G_1 = 1.0$, $G_2 = 10.0$.
  - Population variance of rewards: $\sigma_{G_1} = \sigma_{G_2} = 0$ (no heterogeneity).
  - Number of stamps for rewards: $\bar{S}_1 = 5$, $\bar{S}_2 = 10$.
  - Price coefficient: $\gamma = -1.0$.
  - Discount factor: $\beta = 0.98$.
- We use gaussian kernel with bandwidth=0.02, and estimate α_1, α_2, G_1, G_2, γ.

Estimation 1: Impact of N (cont'd)

Parameter estimates

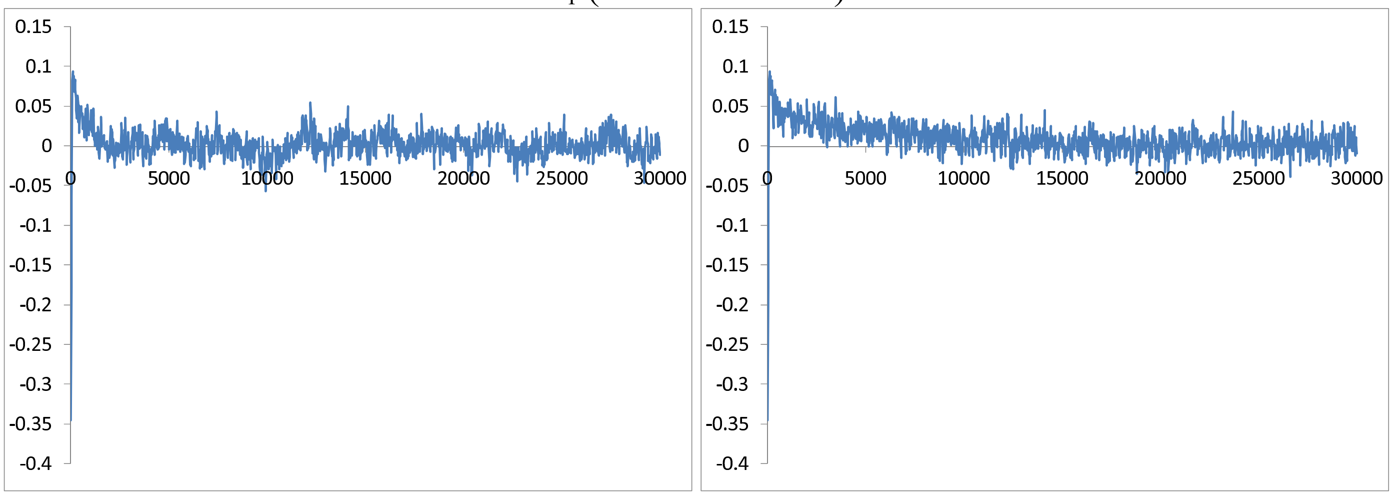
		N=100		N=1000	
parameter	TRUE	mean	sd	mean	sd
α_1 (intercept for store 1)	0.0	-0.049	0.020	-0.061	0.020
α_2 (intercept for store 2)	0.0	0.032	0.019	0.022	0.019
G_1 (reward for store 1)	1.0	1.234	0.034	1.246	0.021
G_2 (reward for store 2)	10.0	9.740	0.063	9.751	0.028
γ (price coefficient)	-1.0	-1.000	0.018	-0.991	0.018

Sample size: 100,000 (1,000 consumers for 100 periods)
Fixed parameters: $\bar{S}_1 = 5$, $\bar{S}_2 = 10$, $\bar{p} = 1.0$, $\sigma_p = 0.3$, $\sigma_{G_j} = 0$ for $j = 1, 2$, $\beta = 0.98$
Tuning parameters: $h = 0.01$ (bandwidth).

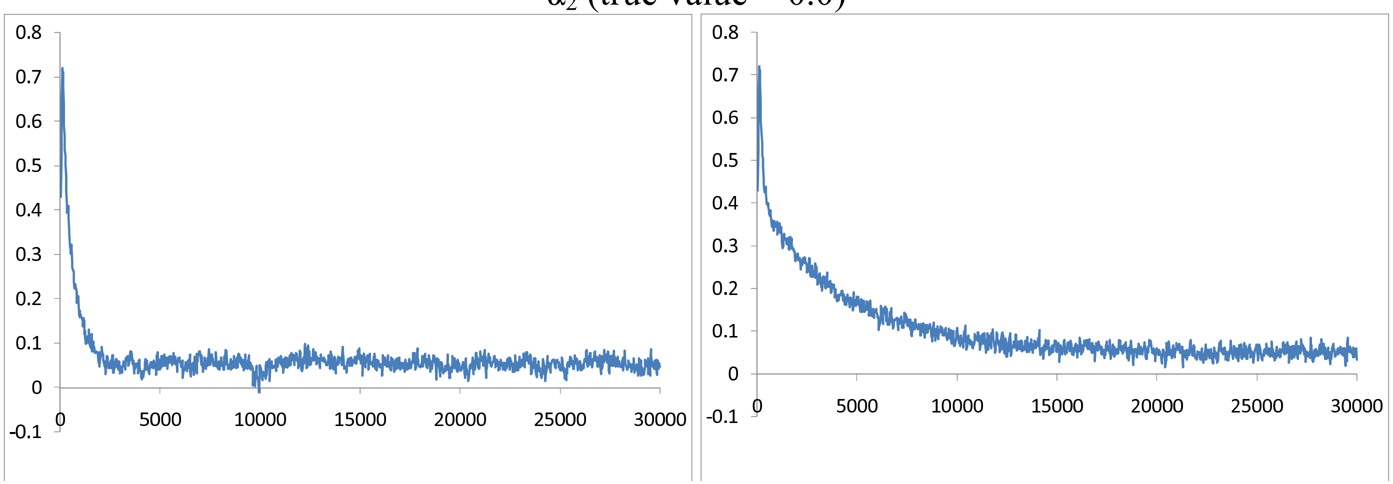
Estimation 1: Impact of N (cont'd)

N = 100 | N = 1000

α_1 (true value = 0.0)



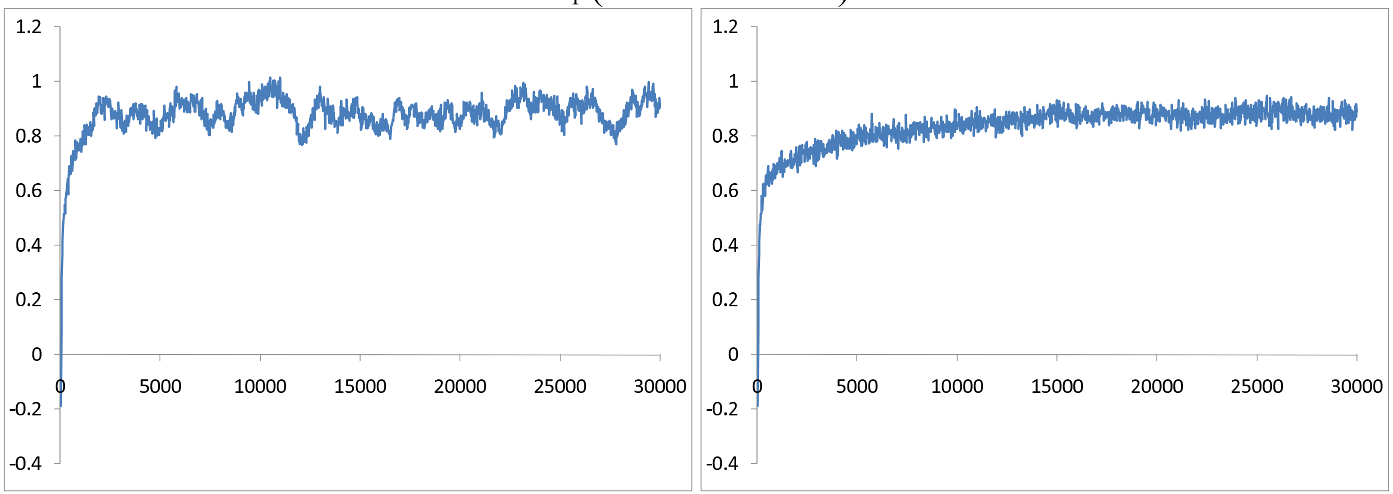
α_2 (true value = 0.0)



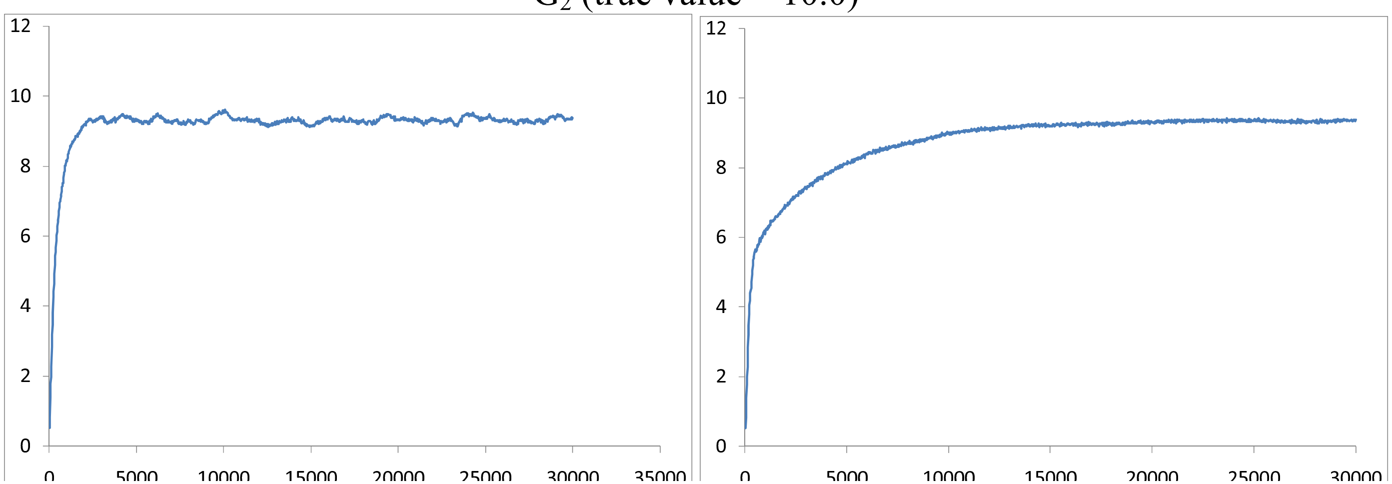
Estimation 1: Impact of N (cont'd)

N = 100 N = 1000

G_1 (true value = 1.0)



G_2 (true value = 10.0)



Estimation 2

- We simulate samples based on the solution of the dynamic model with and without unobserved consumer heterogeneity.
- True parameters of the simulation
  - Store intercepts: $\alpha_1 = \alpha_2 = 0$.
  - Population mean of rewards: $G_1 = 1.0$, $G_2 = 5.0$.
  - Population variance of rewards: $\sigma_{G_1} = 0$, and $\sigma_{G_2} = 0$ (no heterogeneity) or 1.0 (heterogeneity).
  - Number of stamps for rewards: $\bar{S}_1 = 2$, $\bar{S}_2 = 4$.
  - Price coefficient: $\gamma = -1.0$.
  - Discount factor: $\beta = 0.6$ or 0.8.

Results: Homogeneous Model

		$\beta = 0.6$		$\beta = 0.8$	
parameter	TRUE	mean	sd	mean	sd
α_1 (intercept for store 1)	0.0	-0.001	0.019	-0.030	0.022
α_2 (intercept for store 2)	0.0	-0.002	0.019	-0.018	0.028
G_1 (reward for store 1)	1.0	0.998	0.017	1.052	0.021
G_2 (reward for store 2)	5.0	5.032	0.048	5.088	0.085
γ (price coefficient)	-1.0	-0.999	0.016	-0.996	0.019
β (discount factor)	0.6/0.8	0.601	0.008	0.800	0.010

Sample size: 100,000 (1,000 consumers for 100 periods)
Fixed parameters: $\bar{S}_1 = 2$, $\bar{S}_2 = 4$, $\bar{p} = 1.0$, $\sigma_p = 0.3$, $\sigma_{G_j} = 0$ for $j = 1, 2$
Tuning parameters: $N = 1,000$ (no. of past pseudo-value functions used for Emax function approximation), $h = 0.01$ (bandwidth).

Results: Heterogeneous Model

		$\beta = 0.6$		$\beta = 0.8$	
parameter	TRUE	mean	sd	mean	sd
α_1 (intercept for store 1)	0.0	-0.005	0.019	-0.022	0.022
α_2 (intercept for store 2)	0.0	0.010	0.021	0.005	0.037
G_1 (reward for store 1)	1.0	1.017	0.017	1.010	0.019
G_2 (reward for store 2)	5.0	5.066	0.065	4.945	0.130
σ_{G2} (sd of G_2)	1.0	1.034	0.046	1.029	0.040
γ (price coefficient)	-1.0	-1.004	0.016	-0.985	0.019
β (discount factor)	0.6/0.8	0.595	0.005	0.798	0.006

Sample size: 100,000 (1,000 consumers for 100 periods)
Fixed parameters: $\bar{S}_1 = 2$, $\bar{S}_2 = 4$, $\bar{p} = 1.0$, $\sigma_p = 0.3$, $\sigma_{G_1} = 0$
Tuning parameters: $N = 1,000$ (no. of past pseudo-value functions used for Emax function approximation), $h = 0.01$ (bandwidth).

Comparison of Computation Time

Average seconds per iteration

	Homogeneous Model			Heterogeneous Model		
algorithm	$\beta = 0.6$	$\beta = 0.8$	$\beta = 0.98$	$\beta = 0.6$	$\beta = 0.8$	$\beta = 0.98$
Full solution based Bayesian	0.782	0.807	1.410	31.526	65.380	613.26
IJC with N=1000	1.071	1.049	1.006	19.300	19.599	18.387

- For the IJC algorithm, every time a candidate parameter value is rejected, we re-computed the likelihood function at previously accepted parameter value, which will be used to evaluate the likelihood ratio in the Metropolis-Hastings step.
- For the full solution based Bayesian algorithm, we set the number of draws for the integration of prices to be 100.
- Note that the average time per iteration for the full-solution based MCMC increases exponentially with β when estimating a model with unobserved heterogeneity, while it remains unchanged for IJC.
- For $\beta = 0.98$ and no. of iterations = 10,000, full-solution based MCMC would run for 70 days, while IJC would run for 2.5 days.

Policy Experiments

- Policy experiments may seem hard to do from the Bayesian perspective.
- Suppose we are interested in the effect of changing θ_i to $\acute{\theta}_i = (1+t)\theta_i$.
- Modify the IJC algorithm as follows:
  - In each iteration r, set the draw of the policy parameter vector as $\acute{\theta}^r_{-i} = \theta^r_{-i}$ and $\acute{\theta}^r_i = (1+t)\theta^r_i$.
  - Need to store $\{\acute{\theta}^{*l}, E_p\tilde{V}^l(., p; \acute{\theta}^{*l})\}_{l=r-N}^{r-1}$ in addition to what we store before.
  - After convergence, use $\{\theta^r\}_{r=L-D-1}^{L}$ to set the simulated policy parameters: $\acute{\theta}^r_i = (1+t)\theta^r_i$. Then use $\{\acute{\theta}^{*l}, E_p\tilde{V}^l(., p; \acute{\theta}^{*l})\}_{l=r-N}^{r-1}$ to construct the corresponding pseudo-value functions and choice probabilities for the simulated policy experiment outcomes.
  - Extra computation time per iteration for the heterogeneous model: 0.9s for $N = 500$, and 1.74s for $N = 1000$.

Continuous State Space

- What if we allow prices to be serially correlated? (e.g., $f(p'|p;\theta_p)$)
- Modify the IJC algorithm above and combine it with the random grid approximation by Rust (1997).
  - In each iteration r, make a draw of prices, $p^r = (p_1^r, p_2^r)$, from $U[\underline{p}, \bar{p}]^2$, and compute the pseudo-value function at p^r.
  - Outputs of the algorithm become (in the homogeneous model)

$$H^r = \{\{\theta^{*l}, p^l, \tilde{V}^l(., p^l; \theta^{*l})\}_{l=r-N}^{r-1}, \rho^{r-1}(\theta^{r-1})\}$$

  - The expected value function at $(s, p; \theta^r)$ is then approximated as

$$\hat{E}^r_{p'}[V(s', p'; \theta^r)|p] = \sum_{n=1}^{N} \tilde{V}^{r-n}(s', p^{r-n}; \theta^{r-n}) \frac{K_h(\theta^r - \theta^{r-n}) f(p^{r-n}|p;\theta_p)}{\sum_{k=1}^{N} K_h(\theta^r - \theta^{r-k}) f(p^{r-k}|p;\theta_p)}.$$

Remarks

- Curse of Dimensionality of parameter space – in this case, the problem can be mitigated because we can control N.
- If the size of data set is large, and the model is well-identified, MCMC draws would concentrate in a small neighborhood of the posterior means – reduce the need to use a large N.
- There are many kernels that one could use. Norets (2008) extends IJC by using "nearest" neighbors, and consider serially correlated error terms.
- One can also weigh the past pseudo-value functions depending how recent they are.

Remarks (cont'd)

- This paper uses an example of dynamic store choice to illustrate how to implement the IJC algorithm.
- IJC is particularly beneficial in estimating a model with unobserved heterogeneity.
- Osborne (2008) applies the IJC method to estimate a consumer learning model.
- Ching et al. (2009) estimate a dynamic model of learning and forgetting.
- Our C and Matlab programs which implement the IJC method are available upon request.
- The paper is available at SSRN: http://ssrn.com/abstract=1398444

Estimating the Discount Factor

- Can the discount factor and preferences be nonparametrically identified? In general, one needs to make functional form assumptions, or rely on exclusion restrictions (Rust 1994, Magnac and Thesmar 2002).
- Rust (2006) argues that "nonparametric assumptions" such as conditional independence and rational expectation are no weaker than functional form assumptions, such as CARA, error is log normal.
- Many people argue that the discount factor is not well-identified and thus set β according to the interest rate (it is not clear which interest rate one should pick).
- With enough structure in the model, the discount factor can be recovered, along with preference parameters. In particular, exclusion restrictions are particularly helpful (e.g., the number of stamps in the reward program model).

Why study stockpiling?

- The applications of choice models in Marketing started by Guadagni and Little (1983) assumes (i) consumers have complete information about brands, (ii) consumers are myopic.
- But many consumers don't know the quality of all brands in a store. Ignoring this could lead to baised inference.
- Keane (1997) points out the importance of modeling consumer heterogeneity and forward-looking behavior.

Stockpiling

- Blattberg, Eppen and Lieberman (1981), Neslin, Henderson and Quelch (1985), Gupta (1988), Chiang (1991), Bell, Chiang and Padmanabhan (1999), Psendorfer (2002), Hendel and Nevo (2006a) find evidence that consumers stockpile and make purchases strategically.
- Very few research explicitly model consumer forward-looking and stockpiling incentives: Erdem, Imai and Keane (2003), Hendel and Nevo (2006b), Osborne (2009), Gordon and Sun (2009).
- Challenge: The size of the state space is huge because there are many possible combinations of brands and quantities.
- Erdem, Imai and Keane (2003) allows for discrete type unobserved heterogeneity, and assume consumers consume their inventories proportionally across brands in each period.

Hendel and Nevo (2006)

- Hendel and Nevo (2006b) reduce the size of the state space by assuming that brand effects only takes place when you buy the product, and the inventory effect is the same across brands.
- Assume that brand choice decisions and quantity decisions can be separated.
- The state space becomes much smaller.
- Advantage: one can incorporate observed heterogeneity in the brand choice stage very easily.
- Disadvantage: it cannot accommodate unobserved heterogeneity.
- Trade-off: observed heterogeneity with low computational costs vs. unobserved heterogeneity (but limited observed heterogeneity) with high computational costs.

Implementation of Hendel and Nevo (cont'd)

- Step 1: Estimate the brand choice models (including brand-size-time fixed effects).
- Step 2: Assume that prices, advertising, coupon availability can be summarized by an inclusive value term, and model it to follow the first-order markov. [This has significantly simplified the state space.]
- Step 3: Estimate the parameters for the utility of consumptions and storage costs by structurally estimate the simplified dynamic problem.

The Price Consideration Model (Ching, Erdem and Keane 2009a)

- Consumers make a weekly decision whether to consider a category.
- This decision is made prior to seeing prices. However, it may depend on inventory, and promotional activities.
- Only after the consumer decides to consider a category does he/she see prices. In this second stage, the consumer decides whether and what brand to buy.
- Consistent with two stylized facts: (i) many consumers cannot recall the prices of grocery products (Dickson and Sawyer 1990, Vanhule and Dreze 2001, Monroe and Lee 1999); (ii) ad/display could serve as price cues (Anderson and Simester 1998, 2002), or they (together with inventory) could turn on "purchase" mindset (Dhar et al. 2007, Xu and Wyer 2007).

Summary of Results from CEK (2009a)

- The PC model dominates both MNL and Nested MNL on likelihood, AIC and BIC criteria.
- Simulation of data from the models reveals that the PC model produces a dramatically better fit to observed inter-purchase spell lengths than do the MNL and NMNL models.
- In particular, the conventional models greatly exaggerate the probability of short spells. For the PC model, this problem is much less severe.

The Price Consideration Model

- Consider a simple case where price is the only covariate.
- Category has J brands. At each t, prior to seeing prices, consumer decides whether to consider the category.
- $P_{Ct} =$ probability consumer considers category in week t. (This may depend on inventory, promotional activity, etc.)
- If he/she considers the category, then consumer looks at prices, and a MNL with a no-purchase option governs choice behavior.

The PC Model (cont'd)

- Utility of purchase brand j at time t:

  $$U_{jt} = \alpha_j - \beta p_{jt} + e_{jt}$$

  where e_{jt} is an extreme value error.

- Let $P_t(j|C)$ denote the probability the consumer chooses brand j at time t, conditional on considering the category: for $j = 1, \ldots, J$,

  $$P_t(j|C) = \frac{\exp(\alpha_j - \beta p_{jt})}{1 + \sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})} \tag{1}$$

The PC Model (cont'd)

- Let option $J+1$ be no-purchase, with utility set to zero.

$$P_t(J+1|C) = \frac{1}{1 + \sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})} \tag{2}$$

- Then, the unconditional choice probability is:

$$P_t(j) = P_{Ct} P_t(j|C), \tag{3}$$

$$P_t(J+1) = P_{Ct} P_t(J+1|C) + (1 - P_{Ct}). \tag{4}$$

Elaborating the PC Model

- The simple PC model in (1)-(4) can be elaborated in obvious ways:

  1) consumer heterogeneity in the brand intercepts, $\{\alpha_j\}$;
  2) state dependence in brand preferences;
  3) letting the category consideration probability P_{Ct} be a function of feature and display indicators, ad exposures, household size and time since last purchase (to proxy for inventories).

- This is what we did.

The PC model (cont'd)

- The Price Consideration Model (PC) provides a middle ground between the extreme price awareness assumptions of conventional models.
  - Conventional models typically make strong assumptions about when consumers see prices, and when they purchase a category.
  - A model without no-purchase option implicitly assumes the consumer only sees prices after he/she has already decided to buy in a category, and the decision to buy is exogenous.
  - A model with no-purchase option assumes that consumers see prices every week.

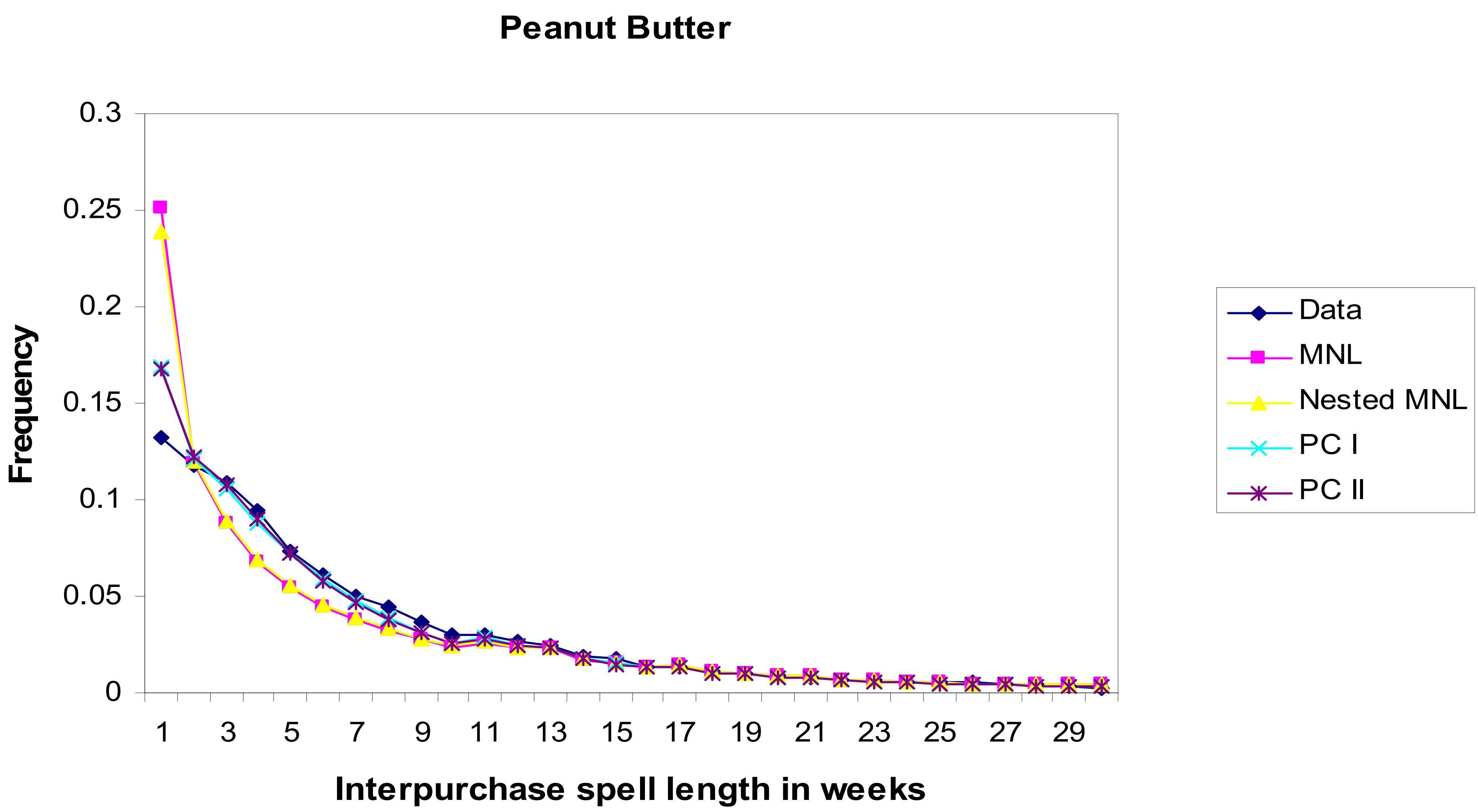
PC Model properties

- In PC, IIA holds among brands, but it does not hold between brands and the no-purchase option.
- But in MNL,IIA holds among brands and the no-purchase option.
- Nested MNL could deviate from IIA. But it can only achieve this by raising the correlation among the extreme value error terms in the second stage, i.e., it forces consumers to switch brands frequently.
- PC can generate this departure without requiring that brands be close substitutes making it more flexible.

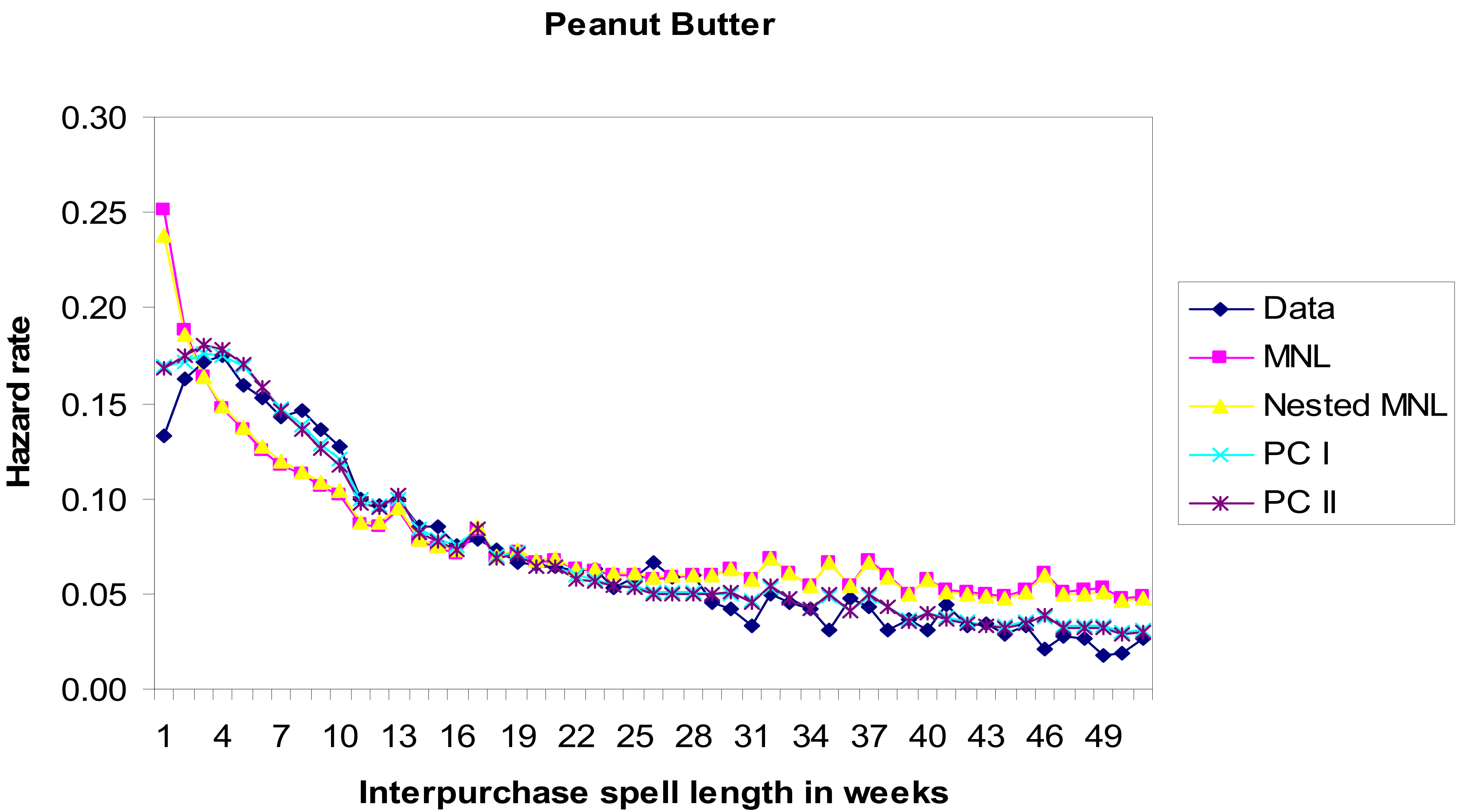
Why does the PC Model Fit Better?

- It generates a more flexible relationship between purchase incidence and brand share price elasticities than conventional models.
- The conventional models greatly overstate the frequency of short inter-purchase spells.
- Because they have difficulty reconciling the observed high sensitivity of brand shares to price with low sensitivity of purchase incidence to price in the period shortly after a purchase (when the inventory is high).
- It is important to distinguish the reason why consumers did not buy. Perhaps, it is much less likely for consumers to walk away once they decide to consider a category (or the purchase mindset is on). If this is the case, identifying what factors can trigger consumers to consider is very crucial from firm's point of view (or for stimulating the economy).

Inter-purchase time distribution



Purchase Hazard



PC Model (cont'd)

- The derivative of the log odds ratio between brands j and k w.r.t. p_{jt}

$$\frac{\partial \ln\left[\frac{P_t(j)}{P_t(k)}\right]}{\partial p_{jt}} = -\beta$$

- The derivative of the log odds ratio between brand j and the no purchase option w.r.t. p_{jt}

$$\frac{\partial \ln\left[\frac{P_t(j)}{P_t(J+1)}\right]}{\partial p_{jt}} = -\beta + \beta \cdot \frac{(1-P_{Ct})\exp(\alpha_j - \beta p_{jt})}{1+(1-P_{Ct})\sum_{k=1}^{J}\exp(\alpha_k - \beta p_{kt})}$$

PC Model (cont'd)

- Nested logit

$$P_t(j|Buy) = \frac{\exp(\alpha_j - \beta p_{jt})}{\sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})}$$

- Define the inclusive value as

$$I_t = \ln\left(\sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})\right)$$

we have

$$P_t(Buy) = \frac{\exp(\rho I_t)}{1 + \exp(\rho I_t)}, \qquad 0 < \rho < 1$$

where $1 - \rho$ is (approximately) the correlation among the extreme value error terms in the second stage.

PC Model (cont'd)

Nested logit (cont'd)

- The unconditional probability of purchase for brand j

$$P_t(j) = P_t(Buy)P_t(j|Buy)$$

This gives

$$\begin{aligned}\frac{P_t(j)}{P_t(J+1)} &= \frac{P_t(j)}{1-P_t(Buy)} = \frac{\frac{\exp(\alpha_j-\beta p_{jt})}{\exp(I_t)}\frac{\exp(\rho I_t)}{1+\exp(\rho I_t)}}{\frac{1}{1+\exp(\rho I_t)}} \\ &= \exp(\alpha_j-\beta p_{jt})\exp([\rho-1]I_t).\end{aligned}$$

Therefore,

$$\frac{\partial \ln\left[\frac{P_t(j)}{P_t(J+1)}\right]}{\partial p_{jt}} = -\beta + \beta\cdot(1-\rho)\cdot\frac{\exp(\alpha_j-\beta p_{jt})}{\sum_{k=1}^{J}\exp(\alpha_k-\beta p_{kt})}$$

PC Model (cont'd)

- Consumer with strong preference for brand 1 (prob. of buying any other brand is negligible).
- PC Model:

$$P_t(1) = P_{Ct} P_t(j|C) \approx P_{Ct} \frac{\exp(\alpha_1 - \beta p_{1t})}{1 + \exp(\alpha_1 - \beta p_{1t})}$$

- Nested logit: $I_t = \alpha_1 - \beta p_{1t}$, and

$$P_t(1) = \frac{\exp(\rho I_t)}{1 + \exp(\rho I_t)} = \frac{\exp(\rho(\alpha_1 - \beta p_{1t}))}{1 + \exp(\rho(\alpha_1 - \beta p_{1t}))}$$

It is not possible to separately identify ρ from α and β

PC Model (cont'd)

When $P_{Ct} = 1$,

$$P_t(1) = P_{Ct}\frac{\exp(\alpha_1 - \beta p_{1t})}{1 + \exp(\alpha_1 - \beta p_{1t})} = \frac{\exp(\alpha_1' - \beta' p_{1t})}{1 + \exp(\alpha_1' - \beta' p_{1t})}$$

The log odds is

$$\ln \frac{P_t(1)}{1 - P_t(1)} = \alpha_1' - \beta' p_{1t}$$

PC model generalizes this to

$$\ln \frac{P_t(1)}{1 - P_t(1)} = \alpha_1 - \beta p_{1t} + \ln P_{Ct} - \ln[1 + (1 - P_{Ct}) \exp(\alpha_1 - \beta p_{1t})]$$

giving

$$\frac{\partial \ln \left[\frac{P_t(1)}{1-P_t(1)}\right]}{\partial p_{1t}} = -\beta + \beta \cdot \frac{(1 - P_{Ct}) \exp(\alpha_1 - \beta p_{1t})}{1 + (1 - P_{Ct}) \exp(\alpha_1 - \beta p_{1t})}$$

PC Model (cont'd)

- NMNL restricts the log odds to be a linear function of price. In the nested logit,

$$\ln \frac{P_t(1)}{1 - P_t(1)} = \rho(\alpha_1 - \beta p_{1t}).$$

- The derivative of the log odds of purchase vs. no purchase w.r.t. a constant shift of the whole price vector

$$\frac{\partial \ln \left[\frac{P_t(Buy)}{P_t(J+1)}\right]}{\partial \bar{P}_t} = -\beta + \beta \cdot \frac{(1 - P_{Ct}) \sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})}{1 + (1 - P_{Ct}) \sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})}$$

while for the nested logit,

$$\frac{\partial \ln \left[\frac{P_t(Buy)}{P_t(J+1)}\right]}{\partial \bar{P}_t} = -\beta\rho$$

Ching, Erdem and Keane (2009b)

- We provide an alternative way to examine how much consumers know about the quality of durable products.
- Key idea: for products depreciate over time and require repeated purchases, observed individuals' inter-purchase spells provide an (objective) measure of quality in terms of durability.
- This is because the higher the durability, the longer a product can last in general.
- We estimate both the subjective perceived brand qualities (based on revealed preference data) and objective brand qualities (based on the brand-specific inter-purchase spells).
- By comparing these two measures, we infer the extent of incomplete information faced by consumers.

An extension of the PC Model

- *Composite* measure of inventory: $I_{it} = I_{it-1} + \pi_j x_{ijt-1} - r_i$, where π_j is the efficiency unit for brand j; x_{ijt} is the quantity purchased of brand j by household i at time t; r_i is the consumption rate.
- In week t, consumer i's consideration probability:

$$P_{it}(C) = \frac{\exp(\gamma_i + Promo_{ct}\gamma_c + log(1 + I_{it}).\gamma_{I,i})}{1 + \exp(\gamma_i + Promo_{ct}\gamma_c + log(1 + I_{it}).\gamma_{I,i})} \quad (5)$$

- γ_i captures consumer heterogeneity.
- $Promo_{ct} = (\sum_{j=1}^{J} Ad_{jt}, \sum_{j=1}^{J} Display_{jt})$, which measure the intensity of advertising and display activities for the category.
- $\gamma_{I,i} = \gamma_0 + Z_{it}.\gamma_z + \epsilon_{Ii}$.
- Z_{it} is a vector of observed characteristics of household i at time t.

Econometric Specification (cont'd)

- Second stage: consumer has decided to consider (but not necessarily buy) in the category.

- Let U_{ijt} denote utility to consumer i of purchasing brand j at time t. For $j = 1, \ldots, J$, let

$$U_{ijt} = \alpha_{ij} + Promo_{jt}\beta + p_{jt}(\varphi_p + Z_{it}\varphi_z + \epsilon_{pi}) + GL(H_{ijt}, \delta) \cdot \lambda + e_{ijt} \quad (6)$$

  - $Promo_{jt}$ is a vector of promotional activites of brand j at time t.
  - $GL(H_{ijt}, \delta)$ is the "brand loyalty" variable defined by Guadagni and Little (1983): $GL_{ij,t} = \delta GL_{ij,t-1} + (1-\delta) d_{ijt-1}$
  - H_{ijt} is the purchase history of household i.

- For $j = J + 1$,

$$U_{i,J+1,t} = e_{i,J+1,t}$$

Why use the market of diaper?

- Pros:
  - Durability - the amount of liquid that it can hold.
  - A market with high turnover of buyers – new buyers keep appearing, and they are more likely to have incomplete information.
  - Repeated purchases happen often – lead to more observations for a fixed period of time (unlike e.g., cars).
  - Only four major brands: Huggies, Pampers, LUVS, Store Brand.
- Cons:
  - Diaper size changes over time (because the baby grows over time).
  - Measures of perceived qualities may reflect style, comfort, etc. Even though they are different from the durability measures, it does not necessarily imply incomplete information.

Data

- Nielsen scanner panel data for LA, Chicago and Atlanta. We observe 819 households, 157 weeks per household from 12/26/1999 to 12/28/2002.
- Choice: when they bought, and which brand they bought.
- Product characteristics: prices, ad, display, diaper size, quantity purchased.
- Household characteristics: female head age, household income, household size, whether has a child < 6 yr old.
- We set the following selection criteria:
  - the household needs to have a member under 6 years old;
  - age of the primary shoppers is below 55;
  - purchase no more than 500 diapers in any given week;
  - purchase at least 5 and less than 50 times;
  - The sample reduces to 346 households.

Summary statistics 1

	No purchase	Huggies	Pampers	LUVS	Store brand	Others	Multiple brands
share (%)	77.89	7.13	5.67	2.4	5.57	0.7	0.65
mean(p_jt)	n.a.	26.91	27.95	22.31	19.87	n.a.	n.a.
mean(ad_jt)	n.a.	0.0223	0.0271	0.0109	0.0123	n.a.	n.a.
mean(display_jt)	n.a.	0.0014	0.0021	0.0016	0.0028	n.a.	n.a.
mean(inter-purch spell)*	n.a.	3.17	2.99	2.9	3.49	3.82	3.13
mean(normalized spell)**	n.a.	0.05	0.048	0.054	0.071	0.093	0.031

* When calculating the "brand-specific" spell, we count a spell belongs to the brand that someone bought at the beginning of the spell.

** Normalized spells are calculated by dividing each spell with the number of diapers bought at the beginning of the spell.

Summary statistics 2

	Sample for PC2	Sample for PC4
#Households	346	131
Average household income*	19.6	19.8
Average household size	4.35	3.5
Average diaper size	3.87	3.77
Average inventory level	130	118.18
Total number of observations	21,173	5,782
Total number of purchases	4,815	1,430

*Household income ranges from 3 to 27: 3=under $5,000; 4=$5,000-7,999; 6=$8,000-9,999; 8=$10,000-11,999;10=$12,000-14,999;11=$15,000-19,999;13=$20000-24999;15=$25000-29999; 16=$30,000-34,999; 17=$35,000-39,999; 18=$40,000-44,999; 19=$45,000-49,999; 21=$50,000-59,999; 23=$60,000-69,999; 26=$70,000-99,999; 27=$100,000 & over.

Results

- PC1 estimates a model without unobserved heterogeneity in the price coefficient and inventory effects.
- PC2 estimates a model with unobserved heterogeneity in the price coefficient and inventory effects, and they could be correlated.
- All specifications allow unobserved heterogeneity in preferences for brand qualities in the 2nd stage.
- Both PC1 and PC2 do not address the initial condition problems (both inventory and GL variables are set to zero in the first observed purchase).
- Set consumption rate, $r_i = 20, \forall i$.

1st Stage Results

	PC 1		PC 2	
	Estimate	s.e.	Estimate	s.e.
Probability of considering a category:				
Pit(C) = exp(Lit)/(1+exp(Lit)) where Lit is determined by the following parameters.				
γ_0	9.433	0.156	8.886	0.288
γ_ad	-0.106	0.500	-0.306	0.831
γ_display	0.200	1.171	0.400	1.553
γ_inventory_0	-2.500		-2.500	
γ_inventory*log(inc)	0.130	0.053	0.213	0.090
γ_inventory*young	-0.003	0.098	0.083	0.425
γ_inventory*hh_size	-0.026	0.033	-0.231	0.039
γ_inventory*diaper_size	0.061	0.032	0.205	0.046
sd(ε_I)			0.072	0.009
π_1(Huggies)	0.641	0.012	0.657	0.004
π_2(Pampers)	0.455	0.021	0.603	0.010
π_3(LUVS)	0.357	0.098	0.414	0.042
π_4(Store Br)	0.575	0.037	0.530	0.027
π_5(Small Br)	0.370	0.003	0.133	0.010
π_6(Multiple Br)	0.171	0.178	0.203	0.059
-2(log-likelihood)	29584.03		29350.32	
AIC	29586.03		29352.32	
BIC	29595.79		29362.08	
#hhold	346		346	

2nd Stage Results

	PC 1		PC 2	
	Estimate	s.e.	Estimate	s.e.
α1 (Huggies)	-2.191	0.158	-1.601	0.170
α2 (Pampers)	-2.446	0.204	-2.221	0.168
α3 (LUVS)	-3.660	0.162	-3.600	0.149
α4 (Store Brand)	-2.579	0.137	-2.280	0.154
Store_brand*LA	0.008	0.242	0.173	0.133
Store_brand*Chicago	-0.304	0.136	-0.418	0.134
β_ad (ad_jt)	0.519	0.212	0.590	0.224
β_d (display_jt)	2.411	0.408	2.321	0.388
Φ_p (p_jt)	-0.054	0.014	-0.077	0.017
Φ_p_inc (p_jt*log(inc))	0.006	0.004	0.017	0.005
Φ_p_young (p_jt*I_young)	0.007	0.005	0.012	0.005
Φ_p_hhsize(p_jt*hh_size)	-0.001	0.001	-0.007	0.001
sd(ε_p)			0.025	0.001
Corr(ε_p, ε_I)			0.127	0.050
State dependence: GL_ijt = δ*GL_ijt-1 + (1- δ)*d_ijt-1				
λ(GL_ijt)	2.458	0.165	2.351	0.186
δ	0.878	0.010	0.880	0.011

Results (cont'd)

- Revealed preference data say: Huggies $\succ$ Pampers $\succ$ Store brand $\succ$ LUVS
- The efficiency units (π_j) also say: Huggies $\succ$ Pampers $\succ$ Store brand $\succ$ LUVS
- Mean (normalized inter-purchase spells) says: Store brand $\succ$ LUVS $\succ$ Huggies $\simeq$ Pampers
- The difference between the rankings of π_j and the normalized spell statistics indicates the importance of controlling for consumer heterogeneity.
- Note that we have not controlled for initial condition problems yet.

Robustness check 1: Initial condition problems

- Try to select households who are "first-time buyers."
- Consider a household with purchase dates: $t_1, t_2, ..., t_N$.
- A household purchase at time t_1 is considered to be its first ever if

$$t_1 > \max_{j=1,2,...,N-1} \{t_{j+1} - t_j\}. \tag{7}$$

- PC3: Same model, but only use "first-time buyers."
- Sample size become 235 households.

Comparison 1: 1st Stage Results

	PC 2		PC 3	
	Estimate	s.e.	Estimate	s.e.
Probability of considering a category:				
Pit(C) = exp(Lit)/(1+exp(Lit)) where Lit is determined by the following parameters.				
γ_0	8.886	0.288	8.115	0.442
γ_ad	-0.306	0.831	-0.308	0.748
γ_display	0.400	1.553	0.402	1.406
γ_inventory_0	-2.500		-2.500	
γ_inventory*log(inc)	0.213	0.090	0.340	0.113
γ_inventory*young	0.083	0.425	0.653	0.202
γ_inventory*hh_size	-0.231	0.039	-0.267	0.067
γ_inventory*diaper_size	0.205	0.046	0.131	0.033
sd(ε_I)	0.072	0.009	0.060	0.003
π_1(Huggies)	0.657	0.004	0.604	0.012
π_2(Pampers)	0.603	0.010	0.576	0.010
π_3(LUVS)	0.414	0.042	0.464	0.011
π_4(Store Br)	0.530	0.027	0.522	0.016
π_5(Small Br)	0.133	0.010	0.233	0.103
π_6(Multiple Br)	0.203	0.059	0.141	0.019
-2(log-likelihood)	29350.32		18360.37	
AIC	29352.32		18362.37	
BIC	29362.08		18372.13	
#hhold	346		235	

Comparison 1: 2nd Stage Results

	PC 2		PC 3	
	Estimate	s.e.	Estimate	s.e.
α1 (Huggies)	-1.601	0.170	-1.213	0.204
α2 (Pampers)	-2.221	0.168	-2.002	0.221
α3 (LUVS)	-3.600	0.149	-3.335	0.192
α4 (Store Brand)	-2.280	0.154	-2.093	0.188
Store_brand*LA	0.173	0.133	0.065	0.162
Store_brand*Chicago	-0.418	0.134	-0.636	0.168
β_ad (ad_jt)	0.590	0.224	0.111	0.274
β_d (display_jt)	2.321	0.388	2.799	0.837
Φ_p (p_jt)	-0.077	0.017	-0.045	0.023
Φ_p_inc (p_jt*log(inc))	0.017	0.005	-0.012	0.006
Φ_p_young (p_jt*I_young)	0.012	0.005	0.015	0.007
Φ_p_hhsize(p_jt*hh_size)	-0.007	0.001	0.005	0.002
sd(ε_p)	0.025	0.001	0.026	0.001
Corr(ε_p, ε_I)	0.127	0.050	-0.518	0.200
State dependence: GL_ijt = δ*GL_ijt-1 + (1- δ)*d_ijt-1				
λ(GL_ijt)	2.351	0.186	2.444	0.196
δ	0.880	0.011	0.876	0.015

Robustness check 2

- α_j's from PC2 and PC3 are not that different.
- Possibly, many "first-time" buyers have experiences with diapers before.
- In our data, "Presence of children and their ages" tells us: (i) under 6 only; (ii) under 6 and 6-12; (iii) under 6 and 13-17; (iv) under 6 and 6-12 and 13-17.
- It's very likely that category (ii)-(iv) have experiences with diapers before.
- PC4: in addition to the criteria in PC3, only use category (i) households
- The number of households in this sample becomes 131.

Comparison 2: 1st Stage Results

	PC 2		PC 4	
	Estimate	s.e.	Estimate	s.e.
Probability of considering a category:				
Pit(C) = exp(Lit)/(1+exp(Lit)) where Lit is determined by the following parameters.				
γ_0	8.886	0.288	7.811	0.678
γ_ad	-0.306	0.831	-0.506	1.144
γ_display	0.400	1.553	0.600	1.404
γ_inventory_0	-2.500		-2.500	
γ_inventory*log(inc)	0.213	0.090	0.606	0.145
γ_inventory*young	0.083	0.425	0.292	0.188
γ_inventory*hh_size	-0.231	0.039	-0.514	0.097
γ_inventory*diaper_size	0.205	0.046	0.221	0.047
sd(ε_I)	0.072	0.009	0.091	0.016
π_1(Huggies)	0.657	0.004	0.602	0.006
π_2(Pampers)	0.603	0.010	0.578	0.011
π_3(LUVS)	0.414	0.042	0.429	0.017
π_4(Store Br)	0.530	0.027	0.521	0.024
π_5(Small Br)	0.133	0.010	0.156	1.688
π_6(Multiple Br)	0.203	0.059	0.373	0.052
-2(log-likelihood)	29350.32		8571.79	
AIC	29352.32		8583.79	
BIC	29362.08		8642.37	
#hhold	346		131	

Comparison 2: 2nd Stage Results

	PC 2		PC 4	
	Estimate	s.e.	Estimate	s.e.
α1 (Huggies)	-1.601	0.170	-0.932	0.308
α2 (Pampers)	-2.221	0.168	-1.803	0.323
α3 (LUVS)	-3.600	0.149	-3.526	0.301
α4 (Store Brand)	-2.280	0.154	-1.472	0.269
Store_brand*LA	0.173	0.133	-0.218	0.311
Store_brand*Chicago	-0.418	0.134	-1.248	0.297
β_ad (ad_jt)	0.590	0.224	-0.220	0.721
β_d (display_jt)	2.321	0.388	3.202	0.776
Φ_p (p_jt)	-0.077	0.017	-0.037	0.034
Φ_p_inc (p_jt*log(inc))	0.017	0.005	-0.026	0.010
Φ_p_young (p_jt*I_young)	0.012	0.005	0.021	0.006
Φ_p_hhsize(p_jt*hh_size)	-0.007	0.001	0.011	0.003
sd(ε_p)	0.025	0.001	0.023	0.003
Corr(ε_p, ε_I)	0.127	0.050	0.084	0.041
State dependence: GL_ijt = δ*GL_ijt-1 + (1- δ)*d_ijt-1				
λ(GL_ijt)	2.351	0.186	2.563	0.487
δ	0.880	0.011	0.904	0.022

PC2 vs. PC4

- The estimates of α_j's are quite different between PC2 and PC4.
- This indicates that consumer learning about brand qualities could be going on.
- The point estimates of π_j remain quite similar across specifications – consistent with the hypothesis that they mainly capture durability.

Extensions

- Right now, we assume consumption rate is exogeneous. It could be a function of I_{it}.
- One could use brand-specific inventory. But this requires us to make another assumption about how consumers choose which brand to use, when they have multiple brands in stock.
- Our framework could also be applied to products where consumption rate is positively correlated quality, e.g., food, soft drinks, snacks, etc. The implication of inter-purchase spell would be the opposite.
  - The shorter the inter-purchase spell, the better the quality.
  - For this type of products, it may be better to model consumption rate to be brand-specific instead of using the efficiency unit approach.

Remarks

- Dynamic stockpiling models are difficult to estimate: Erdem, Imai and Keane (2003) vs. Hendel and Nevo (2006)
- Hendel and Nevo (2006) offers a significant computational advantage by assuming: (i) no unobserved heterogeneity; (ii) no brand-specific inventory effects.
- One could use our framework to check whether these assumptions hold in the data or not.
- Before estimating a learning model, our approach offers a relatively low cost check on the extent of the incomplete information.

Technology Adoption and Switching Costs

- Adoption of a new technology or product often times generate a flow of benefits over time.
- There could also be uncertainty about how the technology might change over time.
- One challenge is that per period benefits are hardly observed. As a result, one cannot measure the adoption costs/switching costs in monetary terms.
- Recent works: Song and Chintagunta (2003), Ryan and Tucker (2007), Yang and Ching (2009).

An example: Yang and Ching (2009)

- Consider adoption decisions of ATM cards for consumers in Italy (mainly a cash economy).
- Once you learn how to use ATM machines, you can enjoy its benefits every period.
- There is uncertainty about how many ATM machines will become available in the future.
- The benefits depend on consumption, income, opportunity costs of time, and age.
- We don't directly observe the per period benefits of adopting ATM cards. But we observe consumers' usage patterns.

Yang and Ching (2009) (cont'd)

- In Italy, checking account pays interests. By observing the cash withdrawal patterns before and after adopting an ATM card and how much cash is needed to finance one's consumption, we can model consumers' cash demand, and use the data to infer the monetary benefits of adopting an ATM card.
- Combining it with the dynamic model of consumer adoption decisions, it is possible to measure the adoption costs in monetary terms.
- The model also provides an alternative explanation about why the elderly are reluctant to adopt new technology.

Yang and Ching (2009) (cont'd)

- Adopt an approach similar to Hendel and Nevo (2006). Assume that the withdrawal decisions are made separated from the adoption decisions and unobserved heterogeneity plays a secondary role.
- One can use Heckman procedure to check whether there is still anything selection issue left.
- Advantage: It significantly reduces the computational costs, it can easily incorporate observed heterogeneity in the 2nd stage (optimal withdrawal decisions).
- Disadvantage: Allowing for unobserved consumer heterogeneity could be costly.

Bayesian Learning

- To address problem (i), academic research in marketing and economics has extended traditional static choice models to explicitly allow for consumer learning.
- Eckstein et al. (1988), Erdem and Keane (1996), Ackerberg (2003), Mehta et al. (2004), Mehta et al. (2008), Crawford and Shum (2005), Anand and Shachar (2005), Erdem, Keane and Sun (2008) explicitly model consumers as bayesian learners, and estimate their models using micro-level data.
- Ching (2000), Narayanan, Manchanda, Chintagunta (2005), Ching and Ishihara (2009a, 2009b), Ching (forthcoming) etc., also model consumers as bayesian learners and allow for learning from others.

Consumption experience signal: $q^e_{ijt} = q_j + \delta_{ijt}$, where $\delta_{ijt} \sim N(0, \sigma^2_\delta)$.

Expected quality: $E[q_j|I(t+1)] = E[q_j|I(t)] + \iota^e_j(t)(\bar{q}^e_{jt} - E[q_j|I(t)])$.

Challenging Issues

- If we assume consumers are forward-looking, we face the curse of dimensionality in the state space.
- the complexity of the likelihood/GMM objective function makes it difficult to search for the global maximum/minimum.
- Some used approximation methods. Some opted to assume consumers are myopic (e.g., Mehta et al. 2004, Mehta et al. 2008, Coscelli and Shum 2004)
- When estimating learning models using product level data, standard BLP estimation method may not apply because there are two unobserved product characteristics: $(E[q_j|I(t)], \xi_{jt})$
  $E[U_{ijt}|I(t)] = E[q_j|I(t)] + \alpha p_j + \xi_{jt} + e_{ijt}$.
- Ching (2000), Ching (2008), Ching and Ishihara (2009a, 2009b) used a pseudo-policy approach to address this problem.

Approximation Methods

- Keane and Wolpin (1994) – interpolation method, needs a terminal period, and solve the model using backward induction. Examples: Erdem and Keane (1996), Crawford and Shum (2005), Sun (2005).
- Rust (1997) – random grid approximation, needs the state variables to evolve stochastically over time, applies to infinite horizon problem. Example: Sriram et al. (forthcoming).
- Ackerberg (2001) – makes use of importance sampling, need to allow all parameters to be heterogeneous. Examples: Ackerberg (2003), Hartmann (2006), Hartmann and Viard (2008), Rossi (2007).
- Imai, Jain and Ching (2009) only applies to infinite horizon problems (which is a good proxy for high frequency data, e.g., weekly data), but can allow for consumer heterogeneity very flexibly.

Conclude (Other Challenges)

- Hyperbolic Discounting: The discount factor for more imminent future is larger than those for more distant future (Loewenstein and Prelec 1992, Laibson 1997), e.g., $f(t) = \beta\delta^t$. Laibson et al. (2007), Fang and Wang (2008) are attacking the identification of hyperbolic discounting. Quantitative Marketing needs to do more research in this area.
- Consumers may not know prices (Dickson and Sawyer 1990, Vanhule and Dreze 2001, Monroe and Lee 1999). Anderson and Simester (1998) (2002) ran a series of field experiments to investigate this. Essentially no dynamic structural models address this issue (see the reduced form models by Ching, Erdem and Keane 2009a, 2009b).
- Should take advantage of new estimation methods – more complicated dynamic structural models can be estimated now!
- Want to know more estimation methods? See Aguirregabiria and Mira (forthcoming).