How Much Do Consumers Know About the Quality of Products? Evidence from the Diaper Market

Andrew Ching, University of Toronto
Tülin Erdem, New York University
Michael Keane, UNSW & ASU

Presentation at the MIO Conference on April 8-9, 2011

Motivation

- How much do consumers know about the quality of products?
- This research question is of fundamental importance in marketing. For instance, if the true quality of my brand is much better than my rivals', but consumers do not know about it, marketing managers may want to run marketing campaign to provide information about quality.
- This is also important for economics. If consumers have little information about the brands, and researchers assume the other way, the welfare analysis can be very far off.

How do we answer this research question?

- A simple direct comparison approach. Brand choice data gives us the perceived quality of brands. If we also know the true quality of brands, we can just compare them.
- So how do we measure the true brand quality?
- Our key idea: for products which depreciate over time and require repeated purchases, observed brand-specific inter-purchase spells provide us with a measure of durability. This is because the higher the durability, the longer a product can last in general.
- Unlike brand choice data (which can measure perceived quality), inter-purchase spell gives us an alternative way to objectively measure a brand's durability.
- IF durability is the main dimension of quality, by comparing these two measures, we could infer the extent of incomplete information faced by consumers.

Objectives

- Formulate this idea in a simple framework, which can measure both perceived brand quality (based on revealed preference, brand choice data) and "true" brand durability (based on inter-purchase spells) simultaneously.
- Provide some evidence that our durability measure captures the "true" durability of a brand.
- We extend the Price Consideration (PC) Model developed in Ching, Erdem and Keane (2009).
- The PC model produces a dramatically better fit to observed hazard rate of short inter-purchase spells than do the MNL and Nested MNL models. In particular, the conventional models greatly exaggerate the probability of short spells. For the PC model, this problem is much less severe.
- But note that this idea can be incorporated into standard choice models (e.g., MNL and Nested MNL) as well.

Road Map

- Model (Ching, Erdem and Keane (2009))
- Data (Scanner Data on the Diaper Market)
- Results
- Extensions
- Conclusion

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

- 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).

PC Model (cont'd)

- In CEK (2009), we are mainly interested in developing an alternative choice model that can capture the price elasticity of demand better.
- We simply used "time since last purchase" as a proxy for inventory.
- It is not brand specific, and we did not control for quantity purchased.
- To use observed inter-purchase spells to measure brand durability, we need to come up with an inventory proxy that can capture brand heterogeneity and control for quantity purchased.
- We also need to control for consumer heterogeneity w.r.t. the impact of inventory.

Inventory variable

- *Composite* measure of inventory:

$$I_{it} = I_{it-1} + \sum_{j=1}^{J} d_{ijt-1} \cdot \pi_j \cdot x_{it-1} - r_i, \tag{1}$$

- π_j is the efficiency unit for brand j.
- $d_{ijt} = 1$ if consumer i chooses brand j.
- x_{it} is the quantity purchased by household i at time t.
- r_i is the consumption rate.

$$r_i = \frac{\sum_{l=1}^{N-1} \sum_{j=1}^{J} d_{ijt_l} \cdot \pi_j \cdot x_{it_l}}{t_N - t_1}, \tag{2}$$

where $t_1, t_2, ..., t_N$ are the observed purchase dates.

Inventory variable (cont'd)

- The ranking of π_j's gives us a relative measure of the durability of brands.
- For quasi-durable products, the higher the value of π_j, the more durable brand j is.
- Note that given our assumption of consumption rate, the more a consumer buys, the higher his consumption rate.
- Our inventory variable is different from Hendel and Nevo (2006), who assume $\pi_l = \pi_k, \forall l, k$.
- It is somewhat related to Erdem, Imai and Keane (2003), who use a quality weighted measure of inventory. But their quality weights are mainly identified from brand choice data. Our efficiency units are mainly identified from inter-purchase spells.

The 1st Stage

- In week t, consumer i's consideration probability:

$$P_{it}(C) = \frac{\exp(\gamma_i + Promo_{ct}\gamma_c + I_{it}.\gamma_{I,i})}{1 + \exp(\gamma_i + Promo_{ct}\gamma_c + I_{it}.\gamma_{I,i})} \tag{3}$$

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

The 2nd Stage

- After a consumer has decided to consider the category, we use multinomial logit with a no-purchase option to capture his brand choice.
- Let U_{ijt} denote utility to consumer i of purchasing brand j at time t. For $j = 1, \ldots, J$, let

$$U_{ijt} = \alpha_{ij} + Promo_{jt}\beta + p_{jt}(\varphi_p + Z_{it}\varphi_z) + GL(H_{ijt}, \delta) \cdot \lambda + e_{ijt}. \quad (4)$$

  - $Promo_{jt}$ 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 $j = 0$ (i.e., no purchase),

$$U_{i0t} = e_{i0t}.$$

The 2nd Stage (cont'd)

- The unconditional choice probability is:

$$P_{it}(j) = P_{it}(C) \cdot P_{it}(j|C), \tag{5}$$

$$P_{it}(0) = P_{it}(C) \cdot P_{it}(0|C) + (1 - P_{it}(C)). \tag{6}$$

Why apply it to the market of diaper?

- Pros:
  - Durability - the amount of liquid that it can hold, the quality of its adhesive tape and its goodness-of-fit.
  - New buyers enter the market exogenously.
  - 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:
  - Babies grow over time – the need for diaper might diminish 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.

A few more words on durability

- In the case of diapers, often times parents need to change diapers when the babies cry. Babies cry because they don't feel comfortable with a wet/dirty diaper, or the fit/softness of a diaper.
- In a sense, the brand specific inter-purchase spells (after controlling for quantity purchased) reflects the preferences of the babies.
- It reflects more basic preferences that comes from sensory, basic needs, etc.
- So one could interpret inter-purchase spells as another type of revealed preference data. It seems to be cleaner than the brand choice data because one's brand choice could be influenced by many things, such as incomplete information.

Empirical implications

- Can we show evidence that the efficiency units are indeed more "objective" and capture the "true" durability of brands?
- Apply our framework to different samples of consumers with varied degree of prior experiences.
- Hypothesis 1: If the estimated efficiency unit captures the true durability, they should be quite stable across these samples.
- Hypothesis 2: The perceived brand qualities estimated using revealed preferences (i.e., the brand intercepts from the 2nd stage of the model) would likely differ across these samples. As we move from inexperienced consumers to experienced consumers, the ranking of the perceived qualities should become closer to that of the efficiency units.
- We will test these two hypotheses. Since the birth of a baby triggers a household to enter the diaper market, we can reasonably separate consumers into experienced vs. inexperienced ones. We consider three samples of households with different degrees of average prior experiences.

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, quantity purchased.
- Household characteristics: female head age, household income, household size, whether has a child $<$ 6 yr old.

Sample 1 (PC1)

Selection Criteria:

- A household needs to have a member under 6 years old.
- Age of the primary shoppers is below 55.
- Purchases no more than 500 diapers in any given week.
- Purchases at least 5 and less than 60 times.
- Households in this sample may enter the market before our sample period starts.
- The sample reduces to 348 households.

Sample 2 (PC2)

Try to select households who are "first-time buyers" from PC1.

- Consider a household with purchase dates: $t_1, t_2, ..., t_N$.
- A household is considered to be "new" if

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

- We are able to control for the initial condition problem.
- Sample size become 234 households. On average, they are active for about 50 weeks.
- This set of households should have less experiences than PC1 on average.
- But it is still possible that they have kids before.

Sample 3 (PC3)

Apply even more stringent selection criteria to PC2.

- "Presence of children and their ages" tells us that a household has kids: (i) < 6 only;(ii) < 6 and 6-12;(iii) < 6 and 13-17;(iv) < 6 and 6-12 and 13-17.
- In PC3, we only use category (i) households.
- The number of households in this sample becomes 101.
- On average, we expect that PC3 consists of households with the least prior experiences.

Summary statistics for household characteristics

Table 1. Summary statistics of household characteristics

	Sample for PC1	Sample for PC2	Sample for PC3
#Households	348	234	101
Average household income*	19.6	19.8	20.24
Average household size	4.32	4.27	3.46
Percentage of households with female head < 30	29.9%	34.2%	44.5%
Percentage of households with female head below college education	76.3%	75.6%	84.2%
Total number of observations	22,032	12,567	5,178
Total number of purchases	4,882	3,083	1,217

*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=$60,00069,999; 26=$70,000-99,999; 27=$100,000 & over.

Summary statistics for product characteristics

PC 1							
	No purchase	Huggies	Pampers	LUVS	Store Brand	Others	Multiple brands
share (%)	76.3	7.57	5.22	2.93	6.54	0.77	0.71
mean(p_jt)	n.a.	27.01	28.03	22.43	19.93	n.a.	n.a.
mean(ad_jt)	n.a.	0.022	0.027	0.011	0.012	n.a.	n.a.
mean(display_jt)	n.a.	0.001	0.003	0.002	0.003	n.a.	n.a.
mean(inter-purch spell)*	n.a.	4.05	4.18	3.91	4.45	5.17	4.25
mean(normalized spell)**	n.a.	0.068	0.072	0.076	0.089	0.135	0.038
PC 3							
	No purchase	Huggies	Pampers	LUVS	Store Brand	Others	Multiple Brands
share (%)	74.6	8.19	8.69	2.67	5.04	0.19	0.68
mean(p_jt)	n.a.	26.39	27.45	21.98	19.53	n.a.	n.a.
mean(ad_jt)	n.a.	0.021	0.026	0.008	0.012	n.a.	n.a.
mean(display_jt)	n.a.	0.001	0.003	0.001	0.003	n.a.	n.a.
mean(inter-purch spell)*	n.a.	4.08	3.77	2.95	4.31	8.10	3.82
mean(normalized spell)**	n.a.	0.062	0.056	0.055	0.076	0.189	0.032

* 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.

Discussion

- Based on the average normalized inter-purchase spells, it appears that Store Brand $\succ$ LUVS $\succ$ Pampers $\succ$ Huggies.
- This is very different from what the market shares suggest.
- But the summary statistics does not take into account consumer heterogeneity in usage patterns.
- By estimating our model, we will take this into account and should be able to obtain a more accurate measure of brand durabilities based on inter-purchase spells.

1st Stage Results

	PC 1		PC 2		PC 3	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
γ_0	-0.190**	0.121	-0.577*	0.131	0.027	0.283
γ_young	-0.289	0.308	0.118	0.234	-0.060	0.318
γ_fedu	0.206**	0.118	0.685*	0.143	0.228	0.283
γ_ad	0.494	0.633	-0.328	0.453	0.741**	0.658
γ_ad*young	-1.026	0.979	-1.791**	0.990	0.998**	0.986
γ_ad*fedu	-0.685	0.701	0.560	0.501	-1.650*	0.740
γ_display	217.3*	10.58	46.08*	0.991	151.50*	3.757
γ_display*young	-11.67*	1.292	-28.82*	1.148	-12.49**	12.23
γ_display*fedu	-210.8*	10.67	-40.67*	0.993	-145.93*	4.239
sd(γ_0)	0.825*	0.044	0.740*	0.036	0.526*	0.114
γ_inventory_0	-4.3e-4	0.002	-4.3e-3	0.002	-1.7e-3	0.006
γ_inventory*log(inc)	-1.0e-3**	5.5e-4	-3.9e-3*	6.3e-4	-7.3e-4	0.002
γ_inventory*young	1.9e-3*	1.3e-4	-0.003**	0.001	2.7e-5	0.002
γ_inventory*hh_size	2.5e-4*	1.5e-4	7.3e-4*	1.7e-4	-2.4e-4**	4.8e-4
π_1(Huggies)	1	n.a.	1	n.a.	1	n.a.
π_2(Pampers)	0.835*	0.100	0.838*	0.041	0.869*	0.077
π_3(LUVS)	0.842*	0.119	0.825*	0.064	0.606*	0.101
π_4(Store Br)	0.863*	0.068	0.842*	0.046	0.947*	0.090
π_5(Small Br)	0.167**	0.147	0.046	0.109	0.117	0.289
π_6(Multiple Br)	0.012	0.128	0.790*	0.102	0.064**	0.059

*5% significance level;
**10% significance level.

Discussion

- Based on the efficiency unit measures (i.e., π_j's), Huggies $\succ$ Pampers $\sim$ Store Brand $\sim$ LUVS in PC1 and PC2. This provides support that the efficiency units capture the true brand durability and is robust across households with different prior experiences.
- π_{LUV} becomes significantly smaller in PC3. But the sample size for PC3 is much smaller.
- Interestingly, $\pi_{Huggies}$ is consistently the highest among all three cases.
- Also, note that the rankings of π_j's are very different from those from the average normalized inter-purchase spells. This indicates that it is important to control for consumer heterogeneity.

2nd Stage Results

	PC 1		PC 2		PC 3	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
α1 (Huggies)	0.066	0.254	0.657**	0.340	1.040*	0.504
α2 (Pampers)	-0.130	0.258	0.712*	0.348	1.501*	0.507
α3 (LUVS)	-1.475*	0.242	-0.985*	0.326	-2.198*	0.624
α4 (Store Brand)	-0.367**	0.200	0.229	0.279	-0.678**	0.460
α5 (small brands)	-2.991*	0.111	-2.396*	0.113	-4.597*	0.339
α6 (multiple brands)	-3.070*	0.113	-2.414*	0.116	-3.344*	0.210
Store_brand*LA	-0.012	0.156	-0.107	0.177	0.469**	0.321
Store_brand*Chicago	-0.360*	0.172	-0.129	0.211	0.226	0.453
λ(GL_ijt)	6.976*	0.299	6.386*	0.297	7.106*	0.685
δ (GL depreciation rate)	0.900*	0.006	0.888*	0.002	0.919*	2.7e-5
β_ad (ad_jt)	0.692*	0.210	0.203	0.315	-0.215	0.579
β_d (display_jt)	0.687*	0.365	1.233*	0.589	1.951*	0.951
Φ_p (p_jt)	-0.090*	0.021	-0.071*	0.026	-0.129*	0.050
Φ_p_inc (p_jt*log(inc))	0.009**	0.006	-0.005	0.007	0.005	0.016
Φ_p_young (p_jt*I_young)	0.033	0.013	0.054*	0.015	0.044*	0.021
Φ_p_hhsize(p_jt*hh_size)	-0.002*	0.001	0.002**	0.002	-0.004**	0.004

Discussion

- PC3: according to the point estimates of the perceived quality (α_j's), Pampers $\succ$ Huggies $\succ$ Store Brand $\succ$ LUVS.
- Let's focus on Pampers and Huggies first. Note that they are fairly similar in terms of advertising expenditures and the number of variety. Thus, we are hoping that their relative perceived quality mainly reflects their relative perceived durability.
- By comparing the trend of the estimated perceived qualities across these three samples, our results suggest that inexperienced consumers (PC3) perceive Pampers to be better on average.
- But as consumers accumulate more experiences (PC3 $\rightarrow$ PC2 $\rightarrow$ PC1), their perception about Pampers and Huggies has started to reverse – this is consistent with our efficiency unit from the first stage.

Discussion (cont'd)

- LUVS and Store Brand remain inferior in terms of perceived qualities across all samples – this could be because they offer fewer varieties than Pampers and Huggies.
- LUVS is worse than Store Brand in perceived qualities – this could be because retailers put their own Store Brand in better shelf spaces.

An illustration

- Suppose that inc=20, hh_size=4, I_young=0 (i.e., female head is older than 30), I_fedu=0 (i.e., female head education is at least college), $I_{it} = 100$ (i.e., 100 Huggies).
- PC1: $Pr(C) = 0.385$; PC3: $Pr(C) = 0.416$.
- Suppose that everything else the same, except that $I_{it} = 50$.
- PC1: $Pr(C) = 0.418$; PC3: $Pr(C) = 0.461$.
- Suppose that we replace 100 Huggies with 100 Store Brand. That is, $I_{it} = 86$.
- PC1: $Pr(C) = 0.394$; PC3: $Pr(C) = 0.421$.

Another illustration

- How much savings would consumers get if they replace Pampers with Store Brand?
- A simple back-of-the-envelope calculation: the average total savings per household would be about \$19.1 in PC2 and \$26.8 in PC3. On average, they are active for about 52 weeks, and bought about 20 diapers per week.
- Even if consumers just focus on the durability dimension, household's deviation from "optimality" are trivial.

Remarks

- Our research question, "How much do consumers know about the quality of products?" is different from "Do consumers do experimentation?".
- Although experimentation is related to learning and incomplete information, consumers can learn from their past experiences without any deliberate experimentation.
- Also, experimentation requires consumers to be forward-looking. This may or may not be the case.
- Even if consumers are myopic and have incomplete information about the vertical differentiation across brands, they could occasionally buy different brands because of temporary price promotion.
- All we assume in Hypothesis 2 is that consumers who have used diapers before, should have more information about the true brand quality.
- This prediction is robust even if consumers are myopic and do not have any incentive to experiment.

Remarks (cont'd)

- Previous research has extended traditional static brand choice models to allow for consumer learning (e.g., Eckstein, Horsky and Raban 1989, Erdem and Keane 1996, Ackerberg 2003, Mehta et al. 2003, Crawford and Shum 2005, Ching 2010a & 2010b, Ching and Ishihara 2010, Chintagunta et al. 2009, Narayanan and Manchanda 2009, etc.)
- With the aid of the structure of learning models, this line of research is able to draw inference on the extent of incomplete information.
- But this type of models are very difficult to estimate.
- Moreover, the exact structure imposed in these models might not be a good proxy for the actual learning behavior – this could lead to misspecification bias.
- Our proposed approach could also provide more insights about how consumers learn.

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 with 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, the efficiency unit should be interpreted differently – the smaller the better.

Concluding Remarks

- We argue that observed consumers' inter-purchase times contain information on the durability of brands.
- We show that the PC model proposed by Ching, Erdem and Keane (2009) provides a unified framework to estimate both perceived quality (more subjective, based on revealed preference) and durability (more objective, based on the actual need).
- We find that our durability measure is relatively stable across samples which vary with consumer prior experiences.
- If consumers only care about durability, they can save some money by switching from national brands to Store Brand. But their loss from "sub-optimal" behavior appears to be quite small.
- Our results are somewhat encouraging. But more research needs to be done to confirm whether our proposed framework provides reliable measures of product durability.

Concluding Remarks (cont'd)

- 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.
- There are other dimensions of quality (e.g., style) that the efficiency unit cannot capture. But our efficiency units can capture "durability." It is an important dimension of quality and difficult to measure from brand choice data alone.