The Price Consideration Model of Brand Choice

Andrew Ching
Rotman School of Management
University of Toronto
Andrew.Ching@Rotman.Utoronto.Ca

Tülin Erdem
Haas School of Business
University of California, Berkeley
erdem@haas.berkeley.edu

Michael Keane
Department of Economics
Yale University
Michael.Keane@yale.edu

May 2005
Revised, March 2006

The Price Consideration Model of Brand Choice

Abstract

The workhorse brand choice models in marketing are the multinomial logit (MNL) and nested multinomial logit (NMNL). These models place strong restrictions on how brand share and purchase incidence price elasticities are related. They predict market shares well, but not inter-purchase spell lengths. In this paper, we propose a new model of brand choice, the "price consideration" (PC) model, that allows more flexibility in the relation between purchase incidence and brand choice elasticities.

In the PC model, consumers do not observe prices in each period. Every week, a consumer decides whether to consider a category. Only then does he/she look at prices and decide whether and what to buy. Using Nielsen scanner data on peanut butter and ketchup, we show the PC model fits much better than MNL or NMNL. Simulations reveal the reason: the PC model provides a vastly superior fit to inter-purchase spells.

1. Introduction

The workhorse brand choice model in quantitative marketing is unquestionably the multinomial logit (MNL). It is often augmented to include a no-purchase option. Occasionally, nested multinomial logit (NMNL) or multinomial probit (MNP) models are used to allow for correlations among the unobserved attributes of the choice alternatives, or the models are extended to allow for consumer taste heterogeneity. There have been rather strong arguments among proponents of different variants of these models.

But in two fundamental ways, all these workhorse brand choice models are more alike than different: (1) they assume essentially static behavior on the part of consumers, in the sense that choices are based only on current (and perhaps past) but not expected future prices, and (2) they all make strong (albeit different) assumptions about when consumers see prices, and when they consider purchasing in a category.

For instance, any brand choice model that does not contain a no-purchase option *is* a purchase timing/incidence model – of a very strong form. This incidence model says that a random and exogenous process[1] determines when a consumer decides to buy in a category, and that the consumer only sees prices after he/she has already decided to buy. On the other hand, standard brand choice models that contain a no-purchase option make an opposite and equally strong assumption: that consumers see prices in <u>every</u> week,[2] and decide whether to purchase in the category based on these weekly price vectors.[3] And, in either case, whether one assumes consumers see prices always or only if they have already decided to buy, standard models assume that consumers then make decisions solely based on current (and perhaps past) prices.

---

[1] By "exogenous" we mean unrelated to prices, promotion, tastes and inventories.
[2] For expositional convenience, throughout the paper we will treat the week as the decision period.
[3] This is also true in a nested logit model, where at the top level consumers decide whether to buy in the category, and in the lower level choose amongst brands. In such a model, the category purchase decision is based on the inclusive value from the lower level of the nest, which is in turn a function of that week's price vector.

In the paper we propose a fundamentally different model of brand choice that we call the "price consideration" or PC model. The difference between this and earlier brand choice models is that, in the PC model, consumers' make a weekly <u>decision</u> about whether to consider a category, and this decision is made <u>prior</u> to seeing any price information. Of course, the decision will depend on inventory, whether the brand is promoted in the media, and so on. Only after the consumer has decided to consider a category does he/she see prices. In this second stage, the consumer decides whether and what brand to buy. Thus, the PC model provides a middle ground between the extreme price awareness assumptions that underlie conventional choice models (i.e., always vs. only when you buy) because in the PC model consumers see prices probabilistically.

We estimate the PC model on Nielsen scanner data for the ketchup and peanut butter categories. We compare the fit of the PC model to both a MNL with a no-purchase option, and a NMNL model with the category purchase decision at the upper level of the nest. All three models incorporate (i) state dependence in brand preferences a la Guadagni and Little (1983), (ii) dependence of the value of no-purchase on duration since last purchase (to capture inventory effects), and (iii) unobserved heterogeneity in brand intercepts.[4]

We find that the PC model produces substantially superior likelihood values to both the MNL and NMNL, and dominates on the 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.

To our knowledge, this severe failure of conventional MNL and NMNL choice models to fit spell distributions has not been previously noted – or, even if it has, it is certainly not widely

---

[4] We assume a multivariate normal distribution of intercepts, so in fact we are estimated what are known as "heterogeneous logit" models (see, e.g., Harris and Keane (1999)).

known. We suspect this is because it is common in the marketing literature to evaluate models based on in-sample and holdout likelihoods, and fit to choice frequencies, while fit to choice dynamics is rarely examined. Since the PC model is as easy to estimate as the conventional models,[5] we conclude it should be viewed as a serious alternative to MNL and NMNL.

2. Problems with Conventional Choice Models, and the PC Alternative

Keane (1997a) argued that conventional MNL and NMNL choice models (with or without no-purchase) could produce severely biased estimates of own and cross-price elasticities of demand, if inventory-planning behavior by consumers is important. To understand his argument, consider this simple example. Say there are two brands, A and B, and that consumers are totally loyal to either one or the other. The following table lists the number who buy A, B or make no-purchase over a 5 week span under two scenarios. First, a no promotion environment:

Week	1	2	3	4	5
A	10	10	10	10	10
B	10	10	10	10	10
No-purchase	80	80	80	80	80

And second a scenario where Brand A is on promotion (say a 20% price cut) in week 2:

Week	1	2	3	4	5
A	10	20	5	5	10
B	10	10	10	10	10
No-purchase	80	70	85	85	80

Now, notice that brand 2's market share goes from 50% to 67% in week 2. Brand B's market share drops to 33%, even though no consumer switches away from B. Thus, a brand choice model without a no-purchase option would conclude that the cross-price elasticity of demand is substantial (i.e., .34/.20 = 1.7), even though it is, in reality, zero.[6]

---

[5] In fact, the PC model may is easier to estimate than the NMNL, as the latter often has a poorly behaved likelihood.
[6] Erdem, Imai and Keane (2003) describe the problem more formally as "dynamic selection bias." Suppose there are several types of consumers, each with a strong preference for one brand. Suppose further that consumers engage in optimal planning behavior. Each period, a consumer considers his/her level of inventories and the vector of brand

Interestingly, including a no-purchase option does not solve the basic problem. Notice there is a post-promotion dip, presumably arising from inventory behavior, that causes the number of people who buy A to drop by 50% in weeks 3 and 4, before returning to normal in week 5. If we took data from the whole 5-week period, we would see that Brand A's average sales when at its "regular" price are 7.5, increasing to 20.0 when it is on promotion. Thus, a conventional static choice model with a no purchase option would conclude the price elasticity of demand is (20.0/7.5 - 1)/(.20) = 8.3. The correct answer is that the short run elasticity is 5 and that the long run elasticity is zero, since all the extra sales come at the expense of future sales.

As Keane (1997a) notes, the exaggeration of the elasticity could be even greater if consumers are able to anticipate future sales. For example, suppose a retailer always puts brand A on sale in week 2 of every 5 week period. As consumers become aware of the pattern, they could concentrate their purchases in week 2, even if their price elasticity of demand were modest. Conventional choice models, however, would infer an enormous elasticity, since a modest 20% price cut causes sales to jump greatly in week two.

An important paper by Sun, Neslin and Srinivasan (2003) shows that the problems with conventional choice models noted by Keane (1997a) are not merely academic. They conduct two experiments. First, they simulate data from a calibrated model where the hypothetical consumers engage in inventory-planning behavior. Thus, Sun et al know (up to simulation noise) the true

---

prices, and decides if it's a good time to buy. In this framework, if brand A has a price cut in week *t*, consumers who buy in week *t* will contain an overrepresentation of the type that prefers A. Thus, in the self-selected subset of consumers who buy in the category in any given week, there is a <u>negative</u> correlation between brand prices and brand preference (i.e., in weeks when price of a brand is low, the sample contains an over-representation of the type of consumers who prefer that brand). This causes the price elasticity of demand to be exaggerated.

Interestingly, this negative correlation between brand prices and brand preference induced by inventory behavior is opposite in sign to the positive correlation that has been the concern of the literature on "endogenous prices." That literature, stemming from Berry et al (1995), deals with a fundamentally different type of data, where prices and sales are aggregated over long periods of time - in contrast to the high frequency (i.e., daily) price variation observed in scanner data. Then, brands with high unobserved quality will tend to be high priced, inducing a <u>positive</u> correlation between brand prices and preferences. This biases price elasticities of demand towards zero. In scanner data, this problem of unobserved brand quality can be dealt with simply by including brand intercepts.

demand elasticities. They estimate a set of conventional choice models on this data; i.e., MNL with and without no-purchase, NMNL with no-purchase. Each model includes consumer taste hetereogeneity and state dependence. The MNL without no-purchase exaggerates cross-price elasticities by about 100%, while MNL and NMNL with no-purchase do so by about 50%. A dynamic structural model of inventory-planning behavior fit to the same data produces accurate elasticity estimates – as expected since it is the "true" model that was used to simulate the data.

Second, Sun et al estimate the same set of models on Nielsen scanner data for ketchup. Strikingly, they obtain the same pattern. Estimated cross-price elasticities from the MNL and NMNL models that include no-purchase are about 50% greater than those implied by the dynamic structural model that accounts for inventory-planning behavior. And those from the MNL without no-purchase are about 100% greater.

Thus, marketing research needs to confront the fact that own and cross-price elasticities of demand are much greater when estimated from pure brand choice models than from models that include a no purchase option. What accounts for this phenomenon? As we've seen, one explanation is the inventory-planning behavior that Keane (1997a) predicted would cause such a problem.[7] In response, several authors, such as Erdem, Imai and Keane (2003), Sun, Neslin and Srinivasan (2003) and Hendel and Nevo (2004), have proposed abandoning conventional static choice models (i.e., MNL and NMNL) in favor of dynamic structural inventory-planning models. Of course, the difficultly with these models is that they are extremely difficult to estimate.

In this paper, we adopt a different course of action. We seek to develop a simple model – no more complex than MNL or NMNL – that does not suffer from the problems noted by Keane

---

[7] Another potential explanation is that brands are viewed as very close substitutes. That would explain why, conditional on purchase in a category, consumers' choice among competing brands appears to be much more sensitive to price than their decision to purchase in the first place. But a well-known fact about consumer behavior is that in many frequently purchased product categories brand loyalty is high (i.e., rates of brand switching are low), implying that brands are not viewed as close substitutes.

(1997) and Sun, Neslin and Srinivasan (2003). The motivation for our approach is two fold:

First, we conjecture that a brand choice model with a more flexible representation for category

consideration/purchase incidence may provide a more accurate reduced form approximation to

consumer's purchase decision rules than the conventional models like MNL and NMNL. Second,

we note there is a simpler way (than estimation of complex dynamic inventory models) to

reconcile the pattern of elasticities across models noted by Sun, Neslin and Srinivasan (2003).

The idea is simply that consumers do not choose to observe brand prices in each period.

More specifically, we propose a two-stage "price consideration" (PC) model in which a

consumer decides, in each period, whether to consider buying in a category. This decision may

be influenced by inventories, advertising, feature and display conditions. If a consumer decides

to consider a category, he/she looks at prices and decides whether and what brand to buy.[8] Such

a model can generate a pattern where consumers are more sensitive to prices when choosing

among brands than when deciding whether to buy in a category. The positive probability that

consumers do not even consider the category (i.e., they do not even see prices) creates a wedge

between the brand choice price elasticity and the purchase incidence price elasticity.

While this model is extremely simple, it has not, as far as we know, been used before in

the marketing literature. It is important to note that the PC model is quite different from a nested

multinomial logit (NMNL) model, where consumers first decide whether to buy in a category

and then, in a second stage, decide which brand to buy. In the first stage of the NMNL, the

decision whether to purchase in the category is a function of the inclusive value from the second

stage, which is, in fact, a price index for the category. Thus, the NMNL model assumes that

---

[8] This two-stage decision process may be (an approximation to) optimal behavior in a version of an inventory-planning model where there is some fixed cost of examining prices, and it is only optimal to examine prices if inventories fall below some critical level, or if consumers observe some advertising/display/feature signal that reduces the cost of observing prices or conveys a signal that prices are likely to be low.

consumers see all prices in all periods. The second stage is different as well, because in our model the second stage includes a no purchase option. That is, if a consumer decides to consider a category, he/she may still decide, upon seeing prices, to choose no-purchase. A NMNL model can certainly generate a pattern that brand choice price elasticities substantially exceed purchase incidence price elasticities, but it would do so by assuming that brands are very similar. As noted above, such an assumption is inconsistent with data on brand switching behavior.

There are a number of ways we test our model against alternative models in the literature. First, we can simply ask whether it fits better than simple MNL and NMNL models that include a no purchase option. Second, our model can be distinguished from the alternative dynamic inventory story for the same phenomenon by looking at categories where inventories are not important. If brand choice price elasticities exceed purchase incidence price elasticities even for non-storable goods, it favors the simple story.

3. The Price Consideration (PC) Model

3.1 Basic Structure and Properties of the PC model, and Comparison to MNL and NMNL

The simplest version of the PC model takes the following form: Consider a category with *J* brands. In each time period *t*, prior to seeing prices, a consumer decides whether to consider the category. Let P_{Ct} denote the probability the consumer considers the category in week *t*. If the consumer decides to consider the category, then he/she looks at prices, and a MNL model with a no-purchase option governs choice behavior. Let $U_{jt} = \alpha_j - \beta p_{jt} + e_{jt}$ denote utility of purchasing brand *j* at time *t*, where e_{jt} is an extreme value error, Then, letting $P_t(j|C)$ denote the probability, conditional on considering the category, that the consumer chooses brand *j* at time *t*, we have:

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

7

Let option $J+1$ be no-purchase. Normalizing the deterministic part of its utility to zero, we have:

$$
(2) \qquad P_t(J+1 \mid C) = \left[1 + \sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt}) \right]^{-1}
$$

Then, the unconditional probability that the consumer buys brand j at time t is:

$$
(3) \qquad P_t(j) = P_{Ct} P_t(j \mid C),
$$

and the unconditional probability of no purchase is:

$$
(4) \qquad P_t(J+1) = P_{Ct} P_t(J+1 \mid C) + (1 - P_{Ct})
$$

The derivative of the log odds ratio between brands j and k with respect to price of j is simply:

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

which is identical to what one obtains in the MNL model, because the P_{Ct} cancel out. However, the derivative of the log odds ratio between brand j and the no purchase option $J+1$ is:

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

This expression lies in the interval $(-\beta, 0)$, approaching $-\beta$ as $P_{Ct} \uparrow 1$. Thus, if $P_{Ct} < 1$, our model implies a greater price elasticity of the brand choice log odds than that for purchase incidence.

Comparison of (5) and (6) reveals the intuition for identification of P_{Ct} in our model. (5) implies that the independence of irrelevant alternatives (IIA) property holds among brands, but (6) shows that it does not hold between brands and the no purchase option. Thus, this type of departure from IIA can be explained by our model with $P_{Ct} < 1$, but not by a standard MNL model with a no purchase option. Thus, the PC model is more flexible than the MNL in terms of how price can affect purchase incidence vs. choice amongst brands.

Next, we examine how price elasticities for purchase incidence vs. brand choice differ between our model and a nested logit model (see Maddala (1983), p. 70). The difference here is more subtle, because a NMNL with a first stage category purchase decision, followed by a second stage brand choice decision, can also generate brand choice price elasticities that exceed purchase incidence price elasticities. What is required is that the coefficient on the inclusive value be less than one in the first stage.

In the nested logit we have:

$$(7) \qquad P_t(j \mid Buy) = \frac{\exp(\alpha_j - \beta p_{jt})}{\sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})}$$

and, defining the inclusive value as:

$$(8) \qquad I_t = \ln\left(\sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})\right)$$

we have:

$$(9) \qquad P_t(Buy) = \frac{\exp(\rho I_t)}{1 + \exp(\rho I_t)}, \qquad 0 < \rho < 1,$$

where $1-\rho$ is (approximately) the correlation among the extreme value error terms in the second stage (see McFadden (1978)). Thus, the unconditional probability of purchase for brand j is:

$$(10) \qquad P_t(j) = P_t(Buy)P_t(j \mid Buy)$$

This gives that:

$$(11) \qquad \frac{P_t(j)}{P_t(J+1)} = \frac{P_t(j)}{1 - P_t(Buy)} = \frac{\frac{\exp(\alpha_j - \beta p_{jt})}{\exp(I_t)} \frac{\exp(\rho I_t)}{1 + \exp(\rho I_t)}}{\frac{1}{1 + \exp(\rho I_t)}} = \exp(\alpha_j - \beta p_{jt})\exp([\rho - 1]I_t)$$

and, therefore:

$$(12) \quad \frac{\partial \ln\left[\dfrac{P_t(j)}{P_t(J+1)}\right]}{\partial p_{jt}} = -\beta + \beta \cdot (1-\rho) \cdot \frac{\exp(\alpha_j - \beta p_{jt})}{\displaystyle\sum_{k=1}^{J} \exp(\alpha_k - \beta p_{kt})}$$

Thus, by setting $\rho<1$, the nested logit also allows the elasticity of the purchase incidence log odds ratio with respect to price to be less than the elasticity of the brand choice log odds ratio.

However, there is a crucial difference between how the NMNL and PC models achieve this divergence of purchase vs. brand choice elasticities. By setting ρ well below 1, the NMNL model forces brands to be close substitutes, implying frequent brand switching by individual consumers. Intuitively, as $\rho \downarrow 0$, the scale of β must increase to maintain any given sensitivity of total category sales to price. But as the scale of β increases, $P_t(j|Buy)$ approaches a step function, equal to 1 if j is the lowest priced brand and zero otherwise. In contrast, the PC model can generate a large divergence between the price sensitivities of purchase incidence and brand choice, without requiring that brands be close substitutes – making it more flexible.[9]

The simple price consideration model described in (1)-(6) can be elaborated in obvious ways. We can allow for consumer heterogeneity in the brand intercepts $\{\alpha_j\}$, and we can also let the category consideration probability P_{Ct} be a function of feature and display indicators, ad exposures, and time since last purchase. If there is a great deal of taste heterogeneity, in the form of very different α vectors across consumers, this model can generate simultaneously the patterns that (i) brand switching is infrequent and that (ii) price elasticities of demand appear to be much greater when estimated from pure brand choice models than from either purchase incidence models or choice models that include a no purchase option.

---

[9] Nevertheless, we were concerned that the PC and NMNL models would be difficult to distinguish empirically. The reason is that we only observe purchases, not whether or not a consumer sees prices/considers a category. And each model has a mechanism to generate brand choice price elasticities that exceed purchase incidence price elasticities. In our empirical application, we found that distinguishing between the two models is not a problem – the PC model clearly fits the data better.

It is also interesting to examine the relation between the NMNL and PC models when there is consumer taste heterogeneity. This is difficult to do in general, but to get some intuition for this case, consider a consumer who so strongly prefers a single brand, let's say brand 1, that his/her probability of buying any other brand is negligible. In most frequently purchased categories, loyal consumers like this appear to comprise a substantial share of the population. For such a consumer, our model implies that:

$$(13) \qquad P_t(1) = P_{Ct}P_t(j \mid C) \approx P_{Ct}\frac{\exp(\alpha_1 - \beta p_{1t})}{1 + \exp(\alpha_1 - \beta p_{1t})}$$

while the nested logit gives:

$$(14) \qquad I_t = \alpha_1 - \beta p_{1t}$$

$$(15) \qquad P_t(1) = \frac{\exp(\rho I_t)}{1 + \exp(\rho I_t)} = \frac{\exp(\rho(\alpha_1 - \beta p_{1t}))}{1 + \exp(\rho(\alpha_1 - \beta p_{1t}))}.$$

It is obvious that, in the nested logit model, it is not possible to separately identify ρ from α and β if all consumers are of this type. The price consideration model does not have the same problem, since there is a loss of generality in setting $P_{Ct}=1$ in equation (3). To see this, suppose that we try to find parameter values α_1' and β' such that:

$$(16) \qquad P_{Ct}\frac{\exp(\alpha_1 - \beta p_{1t})}{1 + \exp(\alpha_1 - \beta p_{1t})} = \frac{\exp(\alpha_1' - \beta' p_{1t})}{1 + \exp(\alpha_1' - \beta' p_{1t})}$$

for all values of p_{1t}. If there are at least 3 observed price levels in the data, then this gives three equations in only two unknowns and a solution will not generally be possible.

Intuitively, a model with a fairly small value of P_{Ct} can generate sales approaching zero for moderately high prices, without simultaneously requiring that purchase probabilities become very large at low prices. In other words, when setting $P_{Ct}=1$, we have that the log odds is a linear function of price:

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

while the PC model generalizes this to:

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

giving:

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

Finally, it is important to note that NMNL also restricts the log odds to be a linear function of price. In the nested logit we have:

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

Thus, in the PC model the log odds ratio is a more flexible function of price than in the NMNL.

This result does not only arise for loyal consumers. Returning to the case where all brands are considered in the second stage, we can derive analogous expressions. We consider the derivative of the log odds of purchase vs. no purchase with respect to a constant shift of the whole price vector, which we denote by $\partial[\cdot]/\partial\overline{P}$. For the price consideration model we have:

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

while for the nested logit we have:

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

Again, the first expression is a nonlinear function of prices if $P_{Ct}<1$, while the later does not depend on prices. Another interesting distinction between the two models is that the former expression approaches $-\beta[1-P_t(Buy|C)]<0$ as $P_{Ct}\downarrow 0$, while the latter approaches 0 as $\rho \downarrow 0$.

In summary, we have described several ways that the PC model is more flexible than either MNL or NMNL. In the empirical section, we'll see that this added flexibility is empirically relevant - the PC model provides a significantly better fit to choice data from two product categories (i.e., ketchup and peanut butter) than do the conventional models.

3.2 Econometric Model Specification for the PC Model

Now we turn to our detailed specification of the PC model. In week *t*, consumer *i*'s probability of considering a category depends on a vector of category promotional activity variables (X_{ct}), household size (mem_i) and time since last purchase ($purch\_gap_i$). Specifically, let

(23) $$P_{it}(C) = \frac{\exp(\gamma_{i0} + X_{ct}\gamma_c + mem_{it}\cdot\gamma_{mem} + purch\_gap_{it}\cdot\gamma_{pg})}{1+\exp(\gamma_{i0} + X_{ct}\gamma_c + mem_{it}\cdot\gamma_{mem} + purch\_gap_{it}\cdot\gamma_{pg})}.$$

We let X_{ct} include indicators for whether <u>any</u> brand in the category is on feature or display, the idea being that these promotional activities may draw consumers' attention to the category. We include household size (mem_i) and time since last purchase ($purch\_gap_{it}$) to capture inventory effects: a longer time since last purchase means the consumer is more likely to have run out of inventory, and hence more likely to consider the category, especially if household size is large.[10] γ_c, γ_{mem} and γ_{pg} are the associated coefficients of X_{ct}, mem_i and $purch\_gap_{it}$, respectively.

Finally, γ_{i0} is a random coefficient that captures unobserved consumer heterogeneity in the likelihood of considering a category. This may arise because some consumers/households have higher usage rates than others. We assume that γ_{i0} is normally distributed.

---

[10] A longer gap may also indicate a consumer is losing interest in the category, as suggested by the empirical death models of Schmittlein, Morrison, Ehrenberg, etc. (see Helsen and Schmittlein (1993) for a review).

Now we turn to the second stage, where a consumer has decided to consider (but not necessarily buy) the category. Let U_{ijt} denote utility to consumer i of purchasing brand j at time t. We allow this utility to depend on observed and unobserved characteristics of the consumer, and interactions among consumer and brand characteristics. Specifically, for $j = 1, ..., J$, let:

$$(24) \qquad U_{ijt} = \alpha_{ij} + X_{jt}\beta + p_{jt}(\varphi_p + Z_{it}\varphi_1) + GL(H_{ijt}, \delta) + e_{ijt}$$

For $j = J+1$ (i.e., the no purchase option), let:

$$(25) \qquad U_{iJ+1t} = e_{i,J+1,t}.$$

The α_{ij} for $j=1,...,J$ are a vector of brand intercepts capture consumer i's tastes for the unobserved attributes of brand j. As utility is measured relative to the no purchase option, an intercept in (25) is not identified. X_{jt} is a vector of observed attributes of brand j at time t, and β is a corresponding vector of utility weights. p_{jt} is price of brand j at time t. We allow for observed heterogeneity in the marginal utility of consumption of the outside good. Thus, Z_{it} is a vector of observed characteristics of consumer i at time t, and the price coefficient is given by $\varphi_p + Z_{it}\varphi_1$.

The term $GL(H_{ijt}, \delta)$ in (24) is the "brand loyalty" or state dependence variable defined by Guadagni and Little (1983) to capture the idea that a consumer who bought a brand frequently in the past is likely to buy it again. Here, H_{ijt} is consumer i's purchase history for brand j prior to time t, and δ is the exponential smoothing parameter; λ is the coefficient mapping GL into the evaluation of utility. Thus, we have $GL_{ijt} = \delta GL_{ij,t-1} + (1-\delta)d_{ij,t-1}$, where $d_{ij,t-1}$ is the indicator function which equals one if consumer i bought brand j at $t-1$, and zero otherwise.[11]

To capture the possible correlation of consumer tastes among brands, the distribution of the vector $\boldsymbol{\alpha_i}$ is assumed to be multivariate normal. Finally, e_{ijt} is an *i.i.d.* extreme value error term that captures the idiosyncratic taste of consumer i for brand j at time t. Thus, (24)-(25) is

---

[11] Thus, GL is an exponentially smoothed weighted average of lagged purchase indicators for brand j by consumer i. Of course, other forms of state dependence are possible, but this is the most common form in marketing.

what is known as a "heterogeneous" MNL model (see, e.g., Harris and Keane (1999)) that allows for both heterogeneity and state dependence in choice behavior.

3.3 Specification of the Alternative MNL and NMNL Models

We also estimate MNL and NMNL models in order to compare their fit to that of the PC model. The specification of MNL is essentially the same as the price consideration model except that (i) the probability of considering a category is assumed to be 1; (ii) the utility of choosing the no-purchase option is assumed to be,

$$(26) \qquad U_{i,J+1,t} = X_{ct}\beta_c + mem_i \cdot \gamma_{mem} + purch\_gap_{it} \cdot \gamma_{pg} + e_{ijt}.$$

Here, we let the value of the no-purchase option depend on the same category feature, display, household size and purchase gap variables that influence the decision to consider the category in the PC model. This is critical in order to allow a fair comparison between the two models. As in (25), the intercept in (26) is normalized to zero for identification.

In NMNL, the utility of choosing the no purchase option is also given by equation (26). Thus, the conditional choice probability in the second stage is given by:

$$(27) \qquad P_t(j \mid Buy, \alpha_{i1}, ..., \alpha_{iJ}) = \frac{\exp(\alpha_{ij} + X_{jt}\beta + p_{jt}(\varphi_p + z_{it}\varphi_1) + GL(H_{ijt}, \delta))}{\sum\limits_{k=1}^{J} \exp(\alpha_{ik} + X_{kt}\beta + p_{kt}(\varphi_p + z_{it}\varphi_1) + GL(H_{ikt}, \delta))}$$

The inclusive value conditioning on $(\alpha_{i1}, ..., \alpha_{iJ})$ now becomes:

$$(28) \qquad I_t \mid (\alpha_{i1}, ..., \alpha_{iJ}) = \ln\left(\sum\limits_{j=1}^{J} \exp(\alpha_{ij} + X_{jt}\beta + p_{jt}(\varphi_p + z_{it}\varphi_1) + GL(H_{ijt}, \delta))\right).$$

The probability of buying in a category is therefore given by,

$$(29) \qquad P_t(Buy \mid \alpha_{i1}, ..., \alpha_{iJ}) = \frac{\exp(\rho I_t)}{\exp(X_{ct}\beta_c + mem_i \cdot \gamma_{mem} + purch\_gap_{it} \cdot \gamma_{pg}) + \exp(\rho I_t)}$$

where $0 < \rho < 1$. Notice that if $\rho = 1$, the NMNL model is equivalent to MNL.

3.4 Estimation Issues

We face an initial conditions problem as we do not observe consumers' choice histories prior to $t=1$ (see Heckman (1981)). This creates two problems: (1) we cannot construct the initial values of GL(H_{ij1}, δ) and *purch_gap$_{i1}$*, and (2), even if we observed these variables they would be correlated with the brand intercepts α_{ij} and the usage rate parameter γ_{i0}. We deal with this problem in two ways: First, we hold out the n weeks of individual choice histories from the estimation sample and use them to impute the initial value of GL(H_{ij1}, δ) and *purch_gap$_{i1}$* for all consumers.[12] Second, we integrate over the joint distribution of GL(H_{ij1}, δ), *purch_gap$_{i1}$*, α_i and of γ_{i0} using a computational convenient procedure proposed by Wooldridge (2003a, b).

In this approach, the distributions of the unobserved heterogeneity terms are allowed to be functions of the initial values of *GL* and *purch_gap*, as follows: In all three models we have:

$$(30) \qquad \alpha_{ij} = \alpha_j + GL(H_{ijt=1}, \delta) \cdot \alpha_{init,GL} + purch\_gap_{i1} \cdot \alpha_{init,pg} + \varepsilon_{ij}$$

where ε_{ij} is multivariate normal with mean zero, and, in the PC model only, for γ_{i0}, we also have:

$$(31) \qquad \gamma_{i0} = \gamma_0 + purch\_gap_{i1} \cdot \gamma_{initial} + \upsilon_i$$

where υ_i is normal with mean zero and variance σ_υ^2.

Next, consider construction of the likelihood function. There is no closed form for the choice probabilities, since $\boldsymbol{\alpha_i}$ is a J·1 multivariate normal vector, and γ_{i0} is normally distributed. Thus, the probability of a consumer's choice history is a *J+1* dimensional integral. For the categories we consider, *J*=4. We have data on many consumers, so many such integrals must be evaluated, rendering traditional numerical integration infeasible. Instead, we use simulated maximum likelihood, in which Monte Carlo methods are used to simulate the integrals (see, e.g., McFadden (1989), Pakes (1986), Keane (1993, 1994)). We use 200 draws for ($\boldsymbol{\alpha_i}$, γ_{i0}).

---

[12] At t=1-n we assume that GL(H_{ij1-n}, δ)=0 and *purch_gap$_{i1-n}$* =1. We set n=10 in peanut butter and 20 in ketchup.

4. Construction of the Data Sets

We use the Nielsen scanner panel data on ketchup and peanut butter for Sioux Falls, SD and Springfield, MO. The sample period is begins in week 25 of 1986 for both categories. It ends in week 34 of 1988 for ketchup, and in week 23 of 1987 for peanut butter. The ketchup category has 3189 households, 114 weeks, 324,795 store visits, and 24,544 purchases, while peanut butter has 7924 households, 51 weeks, 258,136 store visits, and 31,165 purchases.[13]

During this period, there were four major brands in ketchup: Heinz, Hunt's, Del Monte, and the Store Brand; four major brands in peanut butter: Skippy, JIF, Peter Pan and the Store Brand. There are also some minor brands with very small market shares, and we dropped households that bought these brands from the sample.[14] As a result, we lose 558 (out of 7924) households in peanut butter, and 101 (out of 3189) households in ketchup. The number of store visits is reduced to 236,351 in peanut butter, and 314,417 in ketchup. To impute the initial value of *GL* and *purch_gap*, we use the first 10 weeks of individuals' choice history for peanut butter, and 20 weeks for ketchup. This reduced the number of store visits in the estimation sample to 175,675 and 259,310 for peanut butter and ketchup, respectively.

Household characteristics included in Z_{it} are household income (inc_i) and household size (mem_i). Attributes of alternatives included in X_{jt} are an indicator for whether the brand is on display ($display_{jt}$), whether it is a featured item ($feature_{jt}$), and a measure of coupon availability ($coupon\_av_{jt}$). Attributes of the category included in X_{ct} are a dummy for whether at least one of the brands is on display (I_{dt}), and a dummy for whether one of the brands is a featured item (I_{ft}).

---

[13] We assume store visit decisions are exogenous (e.g., a no-visit week arises if the family leaves Sioux Falls on a vacation, which we do not model), and drop weeks when a consumer did not visit a store. If we instead treat these as no-purchase weeks it has little effect on the results, since households visit a store in the large majority of weeks.

[14] If we combine these small brands into "other" their combined market share for is only 0.55% and 0.06% in peanut butter and ketchup, respectively. Since these brands are rarely purchased, their price and promotional activity information is generally missing, given how these variables are constructed in the Nielsen data (see Erdem, Keane and Sun (1999) for a discussion). To avoid these problems, we decide to drop these small brands from the model, and to drop households who have ever chosen "other" in their choice histories.

A difficulty arises in forming the price variable because we model only purchase timing and brand choice, but not quantity choice. Yet each brand offers more than one package size, and price per ounce varies across package sizes. We need to have the price variable be on an equal footing across brands and weeks.[15] Thus, we decided to always use the price of the most common size package when estimating our model - 32 oz for ketchup, 18oz for peanut butter. Admittedly, this introduces some measurement error into prices, but the problem cannot be resolved by a different construction of the price variable – only by introducing quantity choice.

Another problem in constructing the price and promotion variables is that in scanner data we only observe the price paid by the consumer for the brand he/she actually bought. Similarly, we only observe whether a brand is on display or feature when the consumer chooses the brand. Therefore, prices, *display$_j$* and *feature$_j$* for other brands and weeks must be inferred. Erdem, Keane and Sun (1999) discuss this "missing prices" problem in detail.

To deal with this problem, we use the algorithm described in Keane (1997b). It works as follows: (1) Sort through all the data for a particular store on a particular day. If a consumer is found who bought a particular brand, then use the marked price he/she faced as the marked price for that brand in that store on that day. (2) If no one bought a particular brand in a particular store on a particular day, use the average marked price in that store in that week to fill in the price. (3) If no one bought a particular brand in a particular store in a particular week, then use the average marked price of the brand in that store over the whole sample period to fill in the price.

---

[15] The obvious alternative is, on purchase occasions, to use the vector of prices for the size the consumer actually bought, converted to price per ounce. But this begs the question of what price vector to use on occasions when the consumer chose not to buy. Suppose that in non-purchase weeks we use the vector of prices for the most common size (e.g., 32oz for ketchup), also converted to price per ounce. To see the problem this approach creates, assume there are just two sizes, 32oz and 64oz. Furthermore, lets say that on some fraction of purchase occasions consumers buy the 64oz, and that its price per ounce tends to be lower. This will introduce a systematic bias whereby the price per ounce vector tends to be lower on purchase occasions that on non-purchase occasions (since the average price on purchase occasions is a weighted average of 32 and 64 oz prices, while that on non-purchase occasions is just the average 32oz price). Hence, the price elasticity of demand is exaggerated. It is to avoid this problem that chose to base the price vector on a common size for both purchase and non-purchase occasions.

The missing *display_j* and *feature_j* variables are inferred in a similar fashion. Of course, the observed *display_j*'s and *feature_j*'s are dummies equal to 0 or 1. However, since we may use weekly average values or store average values to fill in the missing *display_j*'s and *feature_j*'s, some of them end up falling between 0 and 1 in our data set.

Next we turn to our coupon variable. Keane (1997b) and Erdem, Keane and Sun (1999) discuss the extremely severe endogeneity problem – leading to extreme upward bias in price elasticities – created by use of price net of redeemed coupons as the price variable. Instead, we use a brand/week specific measure of coupon availability, *coupon_av_{jt}*. Basically, this variable measures the average level of coupon redemption for a particular brand in a particular week. The algorithm for constructing it, which is rather involved, is also described in Keane (1997b).

A final problem is that a small percentage of observations have unreasonably high prices, presumably due to measurement/coding errors. So we created a maximum "plausible" price in each category – $3 for peanut butter and $2 ketchup (in 1985-88 nominal dollars) – and replaced prices above this maximum with the mean observed price of the brand. This procedure affected 2% of the observations for peanut butter, and 1.4% of the observations for ketchup.

Summary statistics for households are given in Table 1, while those for brands are in Table 2. Notice that the two categories are rather different. In peanut butter, the fours brands have similar market shares (ranging from 29.9% for Skippy to 19.8% for Peter Pan), while in ketchup Heinz is dominant (with a market share of 63.1%). Households buy peanut butter and ketchup on 11.66% and 7.35% of shopping occasions, respectively. These figures give face validity to the idea of the PC model: when purchase frequency in a category is this low, it seems unlikely that households would check on peanut butter and ketchup prices in every week.[16]

---

[16] We would argue that the PC model is still appropriate for certain categories like milk, which certain households (e.g., those with young children) buy very frequently. This can be captured by heterogeneity in γ_{i0}.

5. Estimation Results

Table 3 presents results obtained by estimating the multinomial logit model (MNL), nested multinomial logit model (NMNL) and price consideration model (PC), using data for the peanut butter category. We estimate two versions of the PC model. In PC-I, decisions to consider the category depend on category feature and display indicators, the purchase-gap, and household size, while in PC-II these factors influence the utility of the no- purchase option as well.[17]

One striking aspect of the results is that coefficient $\alpha_{init,GL}$ from equation 30, that captures how the pre-sample initialized values of the GL variables are related to the brand intercepts, is very large and significant in all the models.[18] Thus, the pre-sample purchase behavior conveys a great deal of information about brand preferences.

Strikingly, there is no evidence for state dependence in the MNL and NMNL models. The coefficient on GL is the wrong sign (and insignificant in NMNL), and the parameter δ is close to one, implying a brand purchase hardly moves GL. The PC-I and PC-II models, on the other hand, imply significant state dependence. The coefficients on GL are 2.7 to 3.2, and the estimates imply a value of $(1-\delta)$, the coefficient on the purchase dummy, of around .05. Thus, e.g., in PC-II a lagged purchase raises the next period utility for buying a brand by .127.

All four models imply similar average price coefficients. The PC-I and PC-II models imply that, for the "average" household the price coefficients are -.421 and -.321, respectively. Thus, e.g., in PC-II, the effect of a lagged brand purchase on current utility from the brand is equivalent to the effect of a .127/.321 = 39 cent price reduction. Mean prices are in the $1.36 to $1.92 range, so this is a substantial state dependence effect.

---

[17] The PC-II model nests MNL. By setting $P_{it}(C) = 1$ for all i and t, PC II becomes MNL. Compared with MNL, PC II has five more parameters, which generate $P_{it}(C)$. One can send $P_{it}(C)$ to 1 by sending the mean γ_{i0} to infinity.
[18] The coefficient $\alpha_{init,GL}$, which captures how the pre-sample purchase gap is related to the brand intercepts, is negative and significant, as expected. That is, households with a longer initial purchase gap tend to like the entire category less. Hence they have generally lower intercepts, making them more likely to chose the no-purchase option.

Part of why the PC models imply substantial state dependence while MNL and NMNL do not is that the coefficient $\alpha_{init,GL}$, while still substantial, is smaller in the PC models by a factor of 4. Thus, the MNL and NMNL models ascribe all the pre-sample differences in purchase behavior to heterogeneity, while the PC models do not.

As we would expect, the display, feature and coupon availability variables all have large and highly significant positive coefficients in the brand specific utility functions in all four models. In the MNL and NMNL models, the utility of the no-purchase option depends *negatively* on the category feature and display indicators, duration since last purchase in the category (*purch-gap*), and on household size. The latter two findings are consistent with inventory behavior. Analogously, in the PC model, the probability of considering the category depends *positively* on the category feature and display indicators, duration since last purchase in the category, and household size. All these effects are highly significant.

It is useful to examine what the PC models imply about the probability of considering the category under different circumstances. Consider a baseline situation with no brand on display or feature, and a household of size 3 that just bought last period (i.e., $purch\_gap_{it} = 1$).[19] Then, the PC-II model estimates imply the probability of considering the peanut butter category is 39.7%, on average.[20] The estimates of γ_f and γ_d are 1.015 and 1.045, respectively. This implies that the consideration probability increases to 75.6% if one or more brands is on display and feature.[21] [Note that, in peanut butter, the category display and feature indicators equal 1 in 4.63% and 9.04% of weeks, respectively]. Finally, the estimate of γ_{pg} is 0.868. This implies that, starting from the baseline, if we increase the purchase gap to 5 weeks, the probability of considering the

---

[19] We also specify that the initial value of *purch_gap*$_{il}$ was 5. This influences the mean of γ_0 as in (31).
[20] We integrate out the unobserved heterogeneity, γ_{i0}, when obtaining the average probability of considering.
[21] That is, when the category display and feature dummies are both set to 1. The predicted effect of feature and display may seem high, but it should be emphasized that this is only the probability that a consumer will check the prices of peanut butter. After checking prices, the consumer may still choose not to buy in the category.

category increases to 90.7%.[22]

The PC-II model implies that, even conditional on having decided to consider the category, the category feature and display indicators, and household size, have significant negative effects on utility of no-purchase. The log-likelihood improvement from PC-I to PC-II is 114 points, while AIC and BIC improve 220 and 179 points, respectively. This suggests feature and display have some influence on brand choice beyond just drawing attention to the category.[23]

Finally, we turn to our main goal, which is to investigate whether the PC model fits the data better than MNL and NMNL. The bottom panel of Table 3 presents log-likelihood, AIC and BIC values for the four models. The likelihood value for NMNL is modestly better than that for MNL (i.e., 50 points). The ρ on the inclusive value is .70, which implies that brands are closer substitutes than suggested by the MNL model (see equation (9), and recall that if $\rho = 1.0$ the models are equivalent). As we discussed in Section 2 and 3.1, NMNL can generate brand choice price elasticities exceeding those for purchase incidence by assuming brands are similar.

However, the log-likelihoods for the PC-I and PC-II models are superior to NMNL by 563.8 points and 677.8 points, respectively. The AIC and BIC produce very similar comparisons. This is to be expected as the models have similar numbers of parameters (i.e., the MNL, NMNL, PCI-I and PC-II models have 28, 29, 31 and 35 parameters, respectively). Thus, the PC models clearly produce better fits to the peanut butter data than the MNL and NMNL models.

Table 4 presents the estimates for the ketchup category. Here the estimate of ρ in the NMNL model is essentially 1, so the MNL and NMNL models are essentially equivalent. Hence, the parameters estimates and likelihood values for the MNL and NMNL are essentially the same.

---

[22] The estimate of γ_{mem} is 0.121, which implies the effect of household size is fairly small. If we reduce household size from 3 to 1, the consideration probability drops from 39.7% to 35.4%.

[23] The Appendix contains variances and correlations of the brand intercepts. We estimate the Cholesky parameters, but report the implied variances and correlations, as they are more informative. In peanut butter, the NMNL model implies larger variances of and larger correlations among the brand specific intercepts than the other models.

The qualitative results for ketchup are quite similar to those for peanut butter.

One small difference is that, in ketchup, the MNL and NMNL models do generate a positive coefficient λ on the GL variable (about 1.3), but it is not statistically significant. Furthermore, the state dependence implied by the point estimates is very weak. The estimates imply a value of $(1-\delta)$, the coefficient on the purchase dummy, of around .005. Thus, a lagged purchase raises the next period utility for buying a brand by only about .007. As the price coefficient at the mean of the data is -.797, the effect of a lagged purchase on the current period utility evaluation for a brand is equivalent to a price cut of less than 1 cent.

As was the case with peanut butter, the PC models again imply much stronger state dependence. The price coefficient for an "average" household is -.964 for PC-I and -.838 for PC-II, which is similar to the values produced by the logit models. But the PC models generate values of $\lambda (1-\delta)$ of about $(4.3)(.028)=.12$, so a lagged purchase is comparable to about a .12/.84 = 14 cent price cut . Mean prices are $0.87 to $1.15, so this is a substantial effect.[24]

Another difference is that category feature and display variables have positive effects on the value of no-purchase in the PC-II model, whereas in peanut butter they had negative effects. Thus, in ketchup, feature and display make a consumer more likely to consider the category, but, conditional on consideration, they make the consumer less likely to actually buy. In peanut butter, they made both consideration and conditional purchase probabilities higher.

To get a sense of how likely a household is to consider ketchup during a store visit, take a situation with no brand on display or feature, and a household of size 3 that bought last period.[25] The estimates of PC-II imply the probability the household considers buying ketchup is 33.4%,

---

[24] The correlations of the brand specific intercepts in the ketchup category show large positive correlations amongst brands 2, 3 and 4 but not with brand 1 (Heinz). This suggests there is basically a Heinz type and a type that regularly buys the other, lower priced, brands.

[25] As in the peanut butter simulation, we assume the initial value of $purch\_gap_{i1}$=5, and integrate over γ_{i0}.

on average. This increases to 90.2% if one or more brands is on display and feature, a larger effect than in peanut butter. [Note that category display and feature indicators equal 1 in 10.3% and 16.0% of weeks, respectively]. Finally, if we increase the purchase gap to 5 weeks, the consideration probability increases from 33.4% to 69.4%, a smaller effect than in peanut butter.

Turning to the issue of model fit, the MNL and NMNL produce essentially identical log-likelihoods, for the reason noted earlier. The PC-I model is superior by 587.4 points, and the PC-II model is superior by 631.3 points. Again, the PC-I and PC-II models have only 2 and 6 more parameters that NMNL, so the AIC and BIC tell a very similar story. Thus, the PC model also produces a clearly better fit to the ketchup data than do the MNL and NMNL models.

6. Model Simulations

6A. Model Simulations: Inter-purchase Spells

We found in section 5 that the PC models fit better than MNL and NMNL based on likelihood and information criteria (AIC, BIC). This fact alone provides no insight into what aspect(s) of the data are better captured by the PC models. Simulation of the models can shed light on this issue. Thus, we simulated several million hypothetical households from each model. To do this, we simulated 2000 artificial households for each household in the data, using the observed characteristics and forcing variables (price, display feature) for that household.

Table 5 reports the fit of each model to unconditional choice frequencies. Here, there is little to choose from amongst the models, as all seem to fit choice frequencies well. Interestingly, in the marketing literature it is unusual to report any aspect of the fit of brand choice models other than the likelihood, information criteria and unconditional choice frequencies.[26] Rarely are

---

[26] For example, in the last five years of *Marketing Science*, the only paper that reports the fit to inter-purchase times for a brand choice model is Sun (2005). Of course, papers that specifically study purchase incidence do report predicted inter-purchase times (see, e.g., Fok, Paap and Franses (2002)), but they abstract from brand choice.

other aspects of fit, like conditional choice frequencies or inter-purchase times, reported. We now turn to such measures.

Table 6 reports choice frequencies conditional on choice in the previous period. Here, the superior fit of the PC models becomes quite apparent. For example, in the data, the probability a household buys Skippy at t given that it bought it at t-1 is 12.5%. The MNL and NMNL put this probability at about 24% to 25% - roughly <u>double</u> the true value. The PC models also predict too high a frequency, 16.6% to 16.9%, but this is only <u>one-third</u> too high, a much smaller error.[27] The story is similar for ketchup. Thus, we see that a key problem with the MNL and NMNL models is that they exaggerate probabilities of very short (i.e., one period) inter-purchase spells.

Figure 1 shows that PC model is far superior to MNL and NMNL in fitting the overall inter-purchase spell distribution. In peanut butter, the logit models roughly double the true frequency of one period spells, as already noted. They get the frequency of two period spells about right, and then underestimate the frequency of 3 through 9 period spells. In contrast, while the PC models slightly exaggerate the probability of one period spells (as already noted), for periods 2 onward their predicted spell frequencies are essentially indistinguishable from those in the data. As we see in the bottom panel of Figure 1, the story is very similar for ketchup.

Figure 2 reports hazard rates, which make clear some additional interesting patterns. Essentially, we see that the MNL and NMNL models completely miss the fact that the hazard rate is low immediately after purchase, rises for a few weeks to a peak at about week 4, and then gradually declines. Instead, they predict that the hazard rate is falling throughout. In contrast, the PC models do capture the basic pattern, although, as noted earlier, they predict too high a purchase hazard in the first period after purchase.

---

[27] Similarly, the probability a household buys any brand in the peanut butter category at t given that it bought it at t-1 is 13.3%. The MNL and NMNL put this probability at about 24% to 25%, while the PC models put it at about 17%.

These results are consistent with our intuition for why the PC model should fit better than the MNL and NMNL logit models, discussed in Sections 1 and 2. The latter models are constrained to have a close link between the sensitivity of brand choice to price/promotion and the sensitivity of category purchase incidence to price/promotion. The PC models break this link, allowing it to provide a much better fit in inter-purchase spells while still fitting brand choices.[28]

Specifically, we see that the PC model does not exaggerate the frequency of short inter-purchase spells nearly so badly as the MNL and NMNL models. Often, deals/promotions last for more than one period. The logit models have difficultly explaining why a consumer would not then be likely to buy in consecutive periods, unless they are rather insensitive to price/promotion in general.[29] The PC model can avoid this problem by saying a consumer is likely to not even consider a category a week after having just bought.

6B. Model Simulations: Promotion Effects

In this section we look at the models' predictions for the impact of promotions. Table 7 reports the impact of a temporary 10% price cut for Skippy, and compares the predictions of the NMNL and PC models.[30] As price cuts are often accompanied by feature/display activity, we also turned on 0.5% of the display dummies and 2.9% of the feature dummies, as described in detail in the footnote to the table.

The similarity of the predictions of the two models is remarkable. The PC-II model predicts a 20.4% increase in Skippy sales in the week of the promotion, while the NMNL model predicts a 21.3% increase. Thus, both models imply short run "price" elasticities of demand of

---

[28] As we noted in Section 3.1, while NMNL can generate a price elasticity of demand for brand choice much larger than for purchase incidence, it must do it within the inclusive value framework. That means NMNL would require that brands be very similar on unobservables. This has implications for other of aspects of behavior, such as brand switching and sensitivity of purchase incidence to prices of different brands. But the PC model can generate a brand choice elasticity greater than the purchase incidence elasticity without constraining substitutability among brands.

[29] In other words, the standard models have difficulty reconciling the observed sensitivity of brand choice to price with the low frequency with which consumers buy again after a short spell.

[30] The MNL and PC-I models make similar predictions, so we do not report them. They are available on request.

26

roughly –2, although it should be stressed this is not a conventional price elasticity because we include the additional promotion activity that typically accompanies a price cut.

Each model implies little brand switching. The NMNL and PC-II models imply that total category sales increase by 5.7% and 5.8%, respectively. The NMNL says 98% of the increase in Skippy sales is from category expansion, while only 2% is switching from other brands. The PC-II model actually says that sales of the other brands increase slightly (about 0.2 to 0.3%), so all the increased Skippy sales is from category expansion. This phenomenon is a spillover effect of promotion that arises in the PC models. If one brand uses features and/or displays, it can attract a consumer's attention to the category, but, when he/she checks out prices, he/she may decide to buy some other brand besides the one that was promoted.

We can also examine the long run effects of the price cut. Their sign is theoretically ambiguous. A price cut may lead to purchase acceleration, where a consumer buys today rather than waiting to buy in a future period (an inventory planning phenomenon). On the other hand, increased sales today can lead to greater future sales through habit persistence/enhanced brand equity, captured by the GL variables.

The NMNL and PC-II make opposite predictions about the direction of these effects, but both agree they are small. According to NMNL, the drop in Skippy sales in weeks 2 through 10 cancels out 3.4% of the increased sales created by the price cut. In contrast, PC-II says Skippy sales increase by roughly 0.52% in weeks 2 through 10. A notable difference between the models is that PC-II implies a post-promotion dip for all brands in weeks 2 to 3, with total category sales falling about in each week 0.4%. The comparable figure for NMNL is only bout 0.16%.

The results for ketchup are a bit different. When we simulate a 10% price cut for Heinz, accompanied by turning on display (feature) with a probability of 3.6% (4.9%), we get much

larger increases in own sales, 35.6% for NMNL and 36.4% for PC-II. Thus, each model implies the own "price" elasticity of demand of roughly -3.6.[31] The PC-II model again implies a larger post-promotion dip, with sales in the category falling about 0.63% in weeks 2 and 3, while NMNL has them falling only about 0.22%. In ketchup, both models say little of the short run increase in Heinz sales is due to brand switching, 1.9% in NMNL and 1.6% in PC-II. Hence, the category spillover effect of promotion that we saw for the PC-II model in the peanut butter category is not so strong in the ketchup category. Finally, the NMNL says 2.7% of the increase in Heinz sales is due to purchase acceleration, while the PC-II says it is essentially zero.

7. Conclusion

In this paper we have proposed a new choice model of brand choice we call the "price consideration" (PC) model, and shown that is it a viable alternative to the workhorse MNL and NMNL models of the marketing literature. The distinguishing feature of the PC model is that it introduces a weekly decision whether or not to consider a product category, and models this decision as depending *only* on non-price factors. Only if a household decides to consider a category does it look at prices and decide whether and what brand to buy.

The PC model can accommodate a more flexible relationship between purchase incidence and brand share price elasticities than conventional MNL and NMNL models. Using data from the peanut butter and ketchup categories, we show that the PC model produces a much better fit, particularly to inter-purchase spells. The standard models greatly overstate the frequency of short 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 inventory is high). Moreover, the PC model is as simple to estimate as standard models.

---

[31] This figure is remarkable close to estimates obtained for the same category by Erdem, Imai and Keane (2003), using a structural inventory model, and Erdem, Keane and Sun (2006), using a learning model.

Table 1. Summary statistics of the data and household characteristics

	Peanut Butter	Ketchup
#Households	7,366	3,088
Max #weeks observed per household	41	94
#Store visits	175,675	259,310
#Purchases	20,478	19,044
Average household income*	5.94	5.99
Average household size	2.84	2.73

*Household income ranges from 1 to 14: 1=less than $5,000; 2=$5,000-9,999; 3=10,000-14,999; ...; 10=$45,000-49,999; 11=$50,000-59,999; 12=$60,000-74,999; 13=$75,000-99,000; 14=$100,000 or more.

Table 2. Summary statistics of product characteristics

Peanut Butter					
Alternative	**No purchase**	**Skippy**	**JIF**	**Peter Pan**	**Store Brand**
#observations	155,197	6,137	4,867	4,061	5,413
share (%)	88.34	3.49	2.77	2.31	3.08
mean(p_{jt})	n.a.	1.842	1.917	1.887	1.366
mean($feature_{jt}$)	n.a.	0.0211	0.0033	0.0038	0.0298
mean($display_{jt}$)	n.a.	0.0036	0.0038	0.0317	0.0085
mean($coupon\_av_{jt}$)	n.a.	0.1030	0.063	0.215	0.0025

Ketchup					
Alternative	**No purchase**	**Heinz**	**Hunt's**	**Del Monte**	**Store Brand**
#observations	240,266	12,042	3,212	1,528	2,262
share (%)	92.65	4.64	1.24	0.59	0.87
mean(p_{jt})	n.a.	1.151	1.146	1.086	0.874
mean($feature_{jt}$)	n.a.	0.0401	0.0433	0.0583	0.0236
mean($display_{jt}$)	n.a.	0.0282	0.0329	0.0266	0.0183
mean($coupon\_av_{jt}$)	n.a.	0.1246	0.0835	0.0240	0.0045

	Peanut Butter	Ketchup
mean(I_{ft})	0.0904	0.1602
mean(I_{dt})	0.0463	0.1027

Table 3. Estimates for the Peanut Butter Category

	MNL		NMNL		PC I		PC II	
	Estimate	s.e.	Estmate	s.e.	Estimate	s.e.	Estimate	s.e.
α_1 (Store Brand)	-5.465	0.148	-6.673	0.212	-3.791	0.078	-4.972	0.158
α_2 (JIF)	-5.309	0.159	-6.474	0.220	-3.591	0.097	-4.804	0.173
α_3 (Peter Pan)	-6.003	0.161	-7.217	0.224	-4.317	0.102	-5.499	0.175
α_4 (Skippy)	-5.115	0.157	-6.251	0.216	-3.387	0.094	-4.596	0.169
$\alpha_{init, gl}$	106.170	6.061	108.471	5.558	23.548	1.702	25.002	2.223
$\alpha_{init, pg}$	-0.045	0.004	-0.082	0.007	-0.041	0.004	-0.037	0.005
β_d (display$_{jt}$)	1.377	0.054	1.471	0.056	1.728	0.048	1.414	0.055
β_f (feature$_{it}$)	2.061	0.039	2.201	0.042	2.259	0.032	2.084	0.040
β_c (coupon_av$_{jt}$)	1.590	0.157	1.986	0.172	1.605	0.164	1.599	0.166
φ_p (p$_{jt}$)	-0.239	0.085	-0.346	0.092	-0.846	0.050	-0.226	0.090
φ_{inc} (p$_{jt}$ ·inc$_i$)	0.017	0.003	0.024	0.003	0.018	0.003	0.018	0.003
φ_{mem} (p$_{jt}$ ·mem$_i$)	-0.047	0.022	-0.071	0.024	0.112	0.006	-0.071	0.023
State dependence: $GL_{ijt} = \delta*GL_{ijt-1} + (1-\delta)*d_{ijt-1}$								
λ(GL$_{ijt}$)	-1.869	0.584	-0.451	0.723	3.179	0.310	2.692	0.320
δ	0.989	0.001	0.989	0.001	0.949	0.004	0.953	0.005
Utility of no purchase:								
β_{fc}(I$_{ft}$)	-0.391	0.030	-0.579	0.034			-0.289	0.036
β_{dc}(I$_{dt}$)	-0.608	0.038	-0.712	0.038			-0.493	0.046
β_{mem}(mem$_i$)	-0.306	0.039	-0.309	0.031			-0.330	0.041
β_{pg}(purch_gap$_{it}$)	-0.023	0.002	-0.019	0.001			0.001	0.002
η			0.702					
ρ (=1/(1+exp(η)))			-0.855	0.122				
Probability of considering a category:								
$P_{it}(C) = exp(L_{it})/(1+exp(L_{it}))$ where:								
$L_{it} = \gamma_{i0} + \gamma_f \cdot I_{ft} + \gamma_d \cdot I_{dt} + \gamma_{pg} \cdot mem_i + \gamma_{pg} \cdot purch\_gap_{it}$, and $\gamma_{i0} = \gamma_0 + \gamma_{initial} \cdot purch\_gap_{i1} + \upsilon_i$								
γ_0					-1.594	0.140	-1.592	0.162
$\gamma_{initial}$					-0.029	0.012	-0.042	0.013
γ_{mem}					0.118	0.030	0.121	0.033
γ_f					1.376	0.102	1.015	0.107
γ_d					1.748	0.183	1.045	0.157
γ_{pg}					0.717	0.035	0.868	0.055
σ_υ					1.163	0.055	1.312	0.088
-log-likelihood	74596.50		74546.882		73983.11		73869.12	
-2(log-likelihood)	149193.00		149093.764		147966.23		147738.23	
AIC	149249.00		149151.764		148028.23		147808.23	
BIC	149539.45		149452.583		148349.79		148171.29	

Table 4. Estimates for the Ketchup Category

	MNL		NMNL		PC I		PC II	
	Estimate	s.e.	Estmate	s.e.	Estimate	s.e.	Estimate	s.e.
α_1 (Heinz)	-3.902	0.170	-3.902	0.176	-2.860	0.084	-3.693	0.181
α_2 (Hunt's)	-5.251	0.171	-5.252	0.176	-4.084	0.087	-4.915	0.182
α_3 (Del Monte)	-6.130	0.173	-6.129	0.181	-4.959	0.095	-5.786	0.187
α_4 (Store Brand)	-5.966	0.166	-5.969	0.174	-4.799	0.082	-5.593	0.182
$\alpha_{init, gl}$	86.105	7.695	86.245	4.269	16.720	1.921	16.327	1.013
$\alpha_{init, pg}$	-0.037	0.002	-0.037	0.002	-0.030	0.002	-0.031	0.003
β_d (display$_{jt}$)	1.097	0.041	1.097	0.042	1.128	0.032	1.137	0.042
β_f (feature$_{it}$)	2.238	0.032	2.238	0.032	2.136	0.024	2.296	0.034
β_c (coupon_av$_{it}$)	1.583	0.105	1.583	0.107	1.835	0.111	1.836	0.111
φ_p (p$_{jt}$)	-1.001	0.148	-1.000	0.154	-1.680	0.080	-0.963	0.154
φ_{inc} (p$_{jt}$ ·inc$_i$)	0.009	0.005	0.009	0.005	0.006	0.005	0.004	0.005
φ_{mem} (p$_{jt}$ ·mem$_i$)	0.055	0.041	0.055	0.042	0.249	0.012	0.037	0.042
State dependence: GL$_{ijt}$ = δ*GL$_{ijt-1}$ + (1- δ)*d$_{ijt-1}$								
λ (GL$_{ijt}$)	1.338	0.800	1.346	0.970	4.235	0.453	4.323	0.438
δ	0.995	0.0005	0.995	0.0003	0.973	0.004	0.972	0.002
Utility of no purchase:								
β_{fc} (I$_{ft}$)	-0.037	0.029	-0.037	0.034			0.244	0.035
β_{dc} (I$_{dt}$)	-0.105	0.034	-0.105	0.038			0.032	0.040
β_{mem} (mem$_i$)	-0.217	0.045	-0.217	0.031			-0.259	0.048
β_{pg} (purch_gap$_{it}$)	-0.011	0.001	-0.011	0.001			-0.002	0.001
η			-10.212	0.632				
ρ (=1/(1+exp(η)))			0.999					
Probability of considering a category: P$_{it}$(C) = exp(L$_{it}$)/(1+exp(L$_{it}$)), where: L$_{it}$ = γ$_{i0}$ + γ$_f$ ·I$_{ft}$ + γ$_d$ ·I$_{dt}$ + γ$_{pg}$ ·mem$_i$ + γ$_{pg}$ · *purch_gap$_{it}$*, and γ$_{i0}$ = γ$_0$ + γ$_{initial}$ · *purch_gap$_{i1}$* + υ$_i$								
γ_0					-1.043	0.163	-0.937	0.157
$\gamma_{initial}$					-0.015	0.008	-0.009	0.008
γ_{mem}					-0.036	0.035	-0.070	0.033
γ_f					1.527	0.107	1.925	0.161
γ_d					1.035	0.141	1.274	0.182
γ_{pg}					0.475	0.028	0.421	0.025
σ_υ					0.866	0.114	0.716	0.107
-log-likelihood	71945.204		71945.196		71357.804		71313.949	
-2(log-likelihood)	143890.408		143890.393		142715.608		142627.899	
AIC	143946.408		143948.393		142777.608		142697.899	
BIC	144244.845		144257.489		143108.021		143070.945	

Table 5. Model Fit: Simulated vs. Actual Brand Choice Frequencies

Peanut Butter					
	No purchase	**Skippy**	**JIF**	**Peter Pan**	**Store Brand**
Data	0.8834	0.0349	0.0277	0.0231	0.0308
MNL	0.8761	0.0371	0.0290	0.0260	0.0318
Nested MNL	0.8768	0.0366	0.0289	0.0260	0.0317
PC I	0.8777	0.0363	0.0284	0.0263	0.0313
PC II	0.8780	0.0362	0.0281	0.0260	0.0316

Ketchup					
	No purchase	**Heinz**	**Hunts**	**Del Monte**	**Store Brand**
Data	0.9266	0.0464	0.0124	0.0059	0.0087
MNL / NMNL	0.9235	0.0494	0.0127	0.0062	0.0082
PC I	0.9235	0.0490	0.0128	0.0062	0.0085
PC II	0.9237	0.0490	0.0128	0.0062	0.0084

Note: The baseline simulation takes the price and promotion variables in the data as given. We simulate 2,000 hypothetical consumer choice histories based on the characteristics of each sample household, as well as the price and promotion history the household faced. Thus, we simulate 7,366 × 2,000 hypothetical choice histories for peanut butter, and 3,088 × 2,000 for ketchup.

Table 6. Choice Probabilities Conditional on Lagged Choice: Model vs. Data

Peanut Butter					
	Choice at time t-1				
	Skippy	**JIF**	**Peter Pan**	**Store Brand**	**Any brand**
Data	0.1252	0.1456	0.1106	0.1453	0.1325
MNL	0.2531	0.2660	0.2339	0.2553	0.2527
Nested MNL	0.2401	0.2517	0.2196	0.2435	0.2394
PC I	0.1687	0.1782	0.1672	0.1664	0.1700
PC II	0.1662	0.1754	0.1625	0.1713	0.1689

Ketchup					
	Choice at time t-1				
	Heinz	**Hunt's**	**Del Monte**	**Store Brand**	**Any brand**
Data	0.0854	0.0805	0.1085	0.0801	0.0858
MNL/Nested MNL	0.1768	0.1461	0.1386	0.1443	0.1651
PC I	0.1154	0.0940	0.0894	0.0952	0.1074
PC II	0.1161	0.0957	0.0925	0.0972	0.1087

Note: See Table 5.

Table 7. Peanut Butter: Effects of Temporary 10% Price Decrease for Skippy

Week	Skippy	JIF	Peter Pan	Store Brand	Total
Nested MNL					
1	21.2841	-0.2710	-0.3132	0.0792	5.7300
2	-0.1458	-0.0969	-0.3133	0.1314	-0.1353
3	-0.3390	0.0184	-0.3173	-0.1583	-0.1862
4	0.1021	0.1595	-0.2367	0.0959	0.0564
5	0.3194	-0.1401	-0.7385	0.1823	-0.0183
6	-0.1932	-0.0766	0.3384	0.2007	0.0237
7	0.1253	-0.1074	0.3088	-0.1853	0.0582
8	-0.2128	0.1744	-0.1635	-0.1546	-0.0967
9	0.2779	-0.2627	-0.4009	-0.2207	-0.1168
10	0.2064	0.3719	0.1010	0.0686	0.1865
Price Consideration II					
	Skippy	JIF	Peter Pan	Store Brand	Total
1	20.4285	0.2198	0.1798	0.3439	5.8103
2	-0.2137	-0.5287	-0.3159	-0.4538	-0.3646
3	-0.5506	-0.3232	-0.3659	-0.4585	-0.4308
4	0.1284	0.6119	-0.1496	0.1304	0.1808
5	0.3795	-0.4063	-0.1184	0.2744	0.0850
6	0.0197	0.1010	0.2580	-0.0761	0.0607
7	0.3277	-0.2566	0.3436	-0.2661	0.0765
8	-0.0332	0.1211	-0.3092	-0.1890	-0.0925
9	0.0997	-0.0453	0.0915	-0.0253	0.0280
10	0.3664	0.3144	-0.3629	0.0722	0.1128

Note: The table reports the percentage change in purchase frequencies for each brand by week, following a temporary 10% price cut for Skippy in week 34. Thus "Week 1" of the simulation is actually week 34 in the data.

We choose week 34 as the base period for simulation because there was no promotion at all (i.e., Skippy's display and feature dummies equal zero for all observations). But price cuts are often accompanied by display and/or feature. To simulate this, we use the following procedure: First, we calculated the frequency with which display and feature are 1 when Skippy is on "deal." These figures were 0.0052 and 0.02895, respectively. Then, in the simulation, we set Skippy's display dummy to in week 34 to 1 with probability 0.0052, and we set its feature dummy to 1 with probability 0.02895 in the simulation.

When we set the brand specific display (feature) dummy to 1, we also adjust the categorical level display (feature) dummy accordingly (setting it to 1 if it had been 0 in the data). In week 34, the average frequency of the categorical display and feature dummies are 0.0631 and 0.0265 in the baseline, respectively; their average frequencies in the simulation are 0.0681 and 0.0548, respectively. Their average frequencies for the whole sample are 0.0469 and 0.0905, respectively.

**Table 8. Ketchup: Effects of Temporary 10% Heinz Price Decrease**

	Nested MNL				
Week	Heinz	Hunt's	Del Monte	Store Brand	Total
1	35.6246	-1.6961	-2.0031	-0.5379	23.0642
2	-0.2775	-1.2825	0.0699	1.3003	-0.2368
3	-0.4627	-0.0037	-0.8016	0.8059	-0.2022
4	0.0645	0.2452	-0.2097	0.9260	0.1177
5	0.0679	0.6737	0.2956	-0.0373	0.1877
6	-0.2361	0.4360	-0.8697	0.0605	-0.1738
7	0.3613	-1.2993	0.9201	0.0127	0.1404
8	-0.1033	0.1021	0.1655	0.3958	0.0233
9	-0.1407	-0.3851	0.3343	0.6866	-0.0271
10	0.2057	0.8462	-0.6548	0.3882	0.2549
	Price Consideration II				
	Heinz	Hunt's	Del Monte	Store Brand	Total
1	36.4323	-1.2495	-1.8488	-0.4993	23.5150
2	-0.5822	-2.0916	-1.0833	1.1292	-0.6591
3	-0.9858	-0.3131	-0.7486	0.3057	-0.6102
4	-0.1513	0.7361	-1.0777	0.6298	-0.1177
5	0.0522	0.1579	-0.1911	0.0713	0.0593
6	-0.1331	0.1108	-0.9187	-0.2845	-0.1586
7	0.6772	-1.1387	0.7906	-0.4079	0.3240
8	0.1858	-0.4994	0.4755	-0.0057	0.1356
9	0.0434	-0.0739	0.2076	0.7233	0.1110
10	0.2487	0.8157	-0.3368	0.1754	0.2843

Note: The table reports the percentage change in purchase frequencies for each brand by week, following a temporary 10% price cut for Heinz in week 44. Thus "Week 1" of the simulation is actually week 34 in the data.

We choose week 44 as the base period for simulation because there was (almost) no promotion at all (i.e., Heinz' display dummies equals zero for all observations, and feature was 1 for only 0.7% of observations). But price cuts are often accompanied by display and/or feature. To simulate this, we use the following procedure: First, we calculated the frequency with which display and feature are 1 when Heinz is on "deal." These figures were 0.0357 and 0.0489, respectively. Then, in the simulation, we set Skippy's display dummy to in week 44 to 1 with probability 0.0357, and we set its feature dummy to 1 with probability 0.0489 in the simulation.

When we set the brand specific display (feature) dummy to 1, we also adjust the categorical level display (feature) dummy accordingly (setting it to 1 if it had been 0 in the data). In week 44, the average frequency of the categorical display and feature dummies in the baseline are 0.0858 and 0.0551 in the baseline, respectively; their average frequencies in the simulation are 0.1184 and 0.1014, respectively. Their average frequencies for the whole sample are 0.1018 and 0.1606, respectively.

Figure 1: Inter-Purchase Time Distribution

Peanut Butter

Ketchup
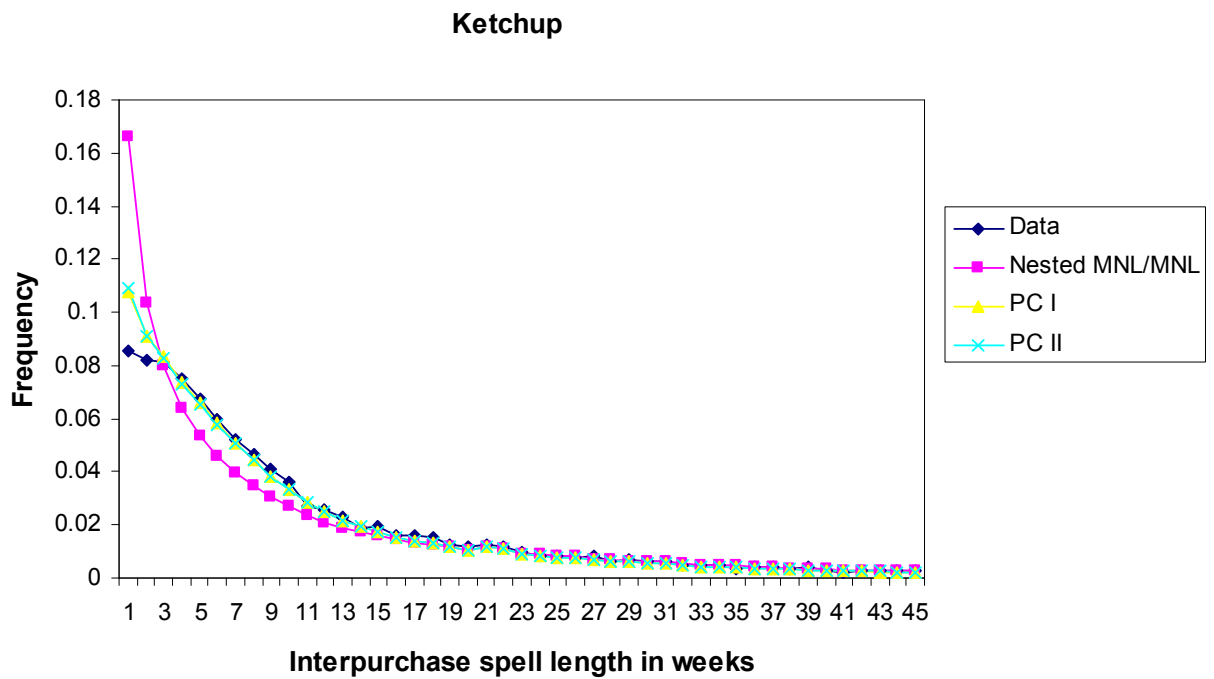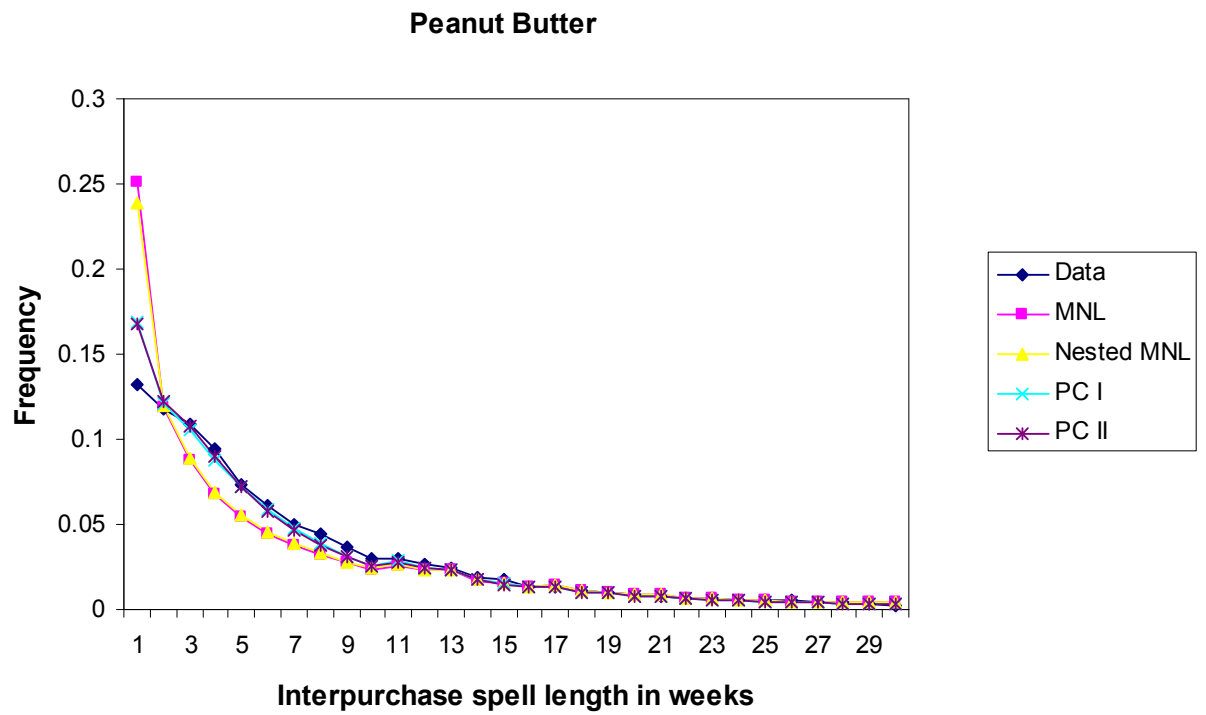
Figure 2. Purchase Hazard

Peanut Butter

Ketchup
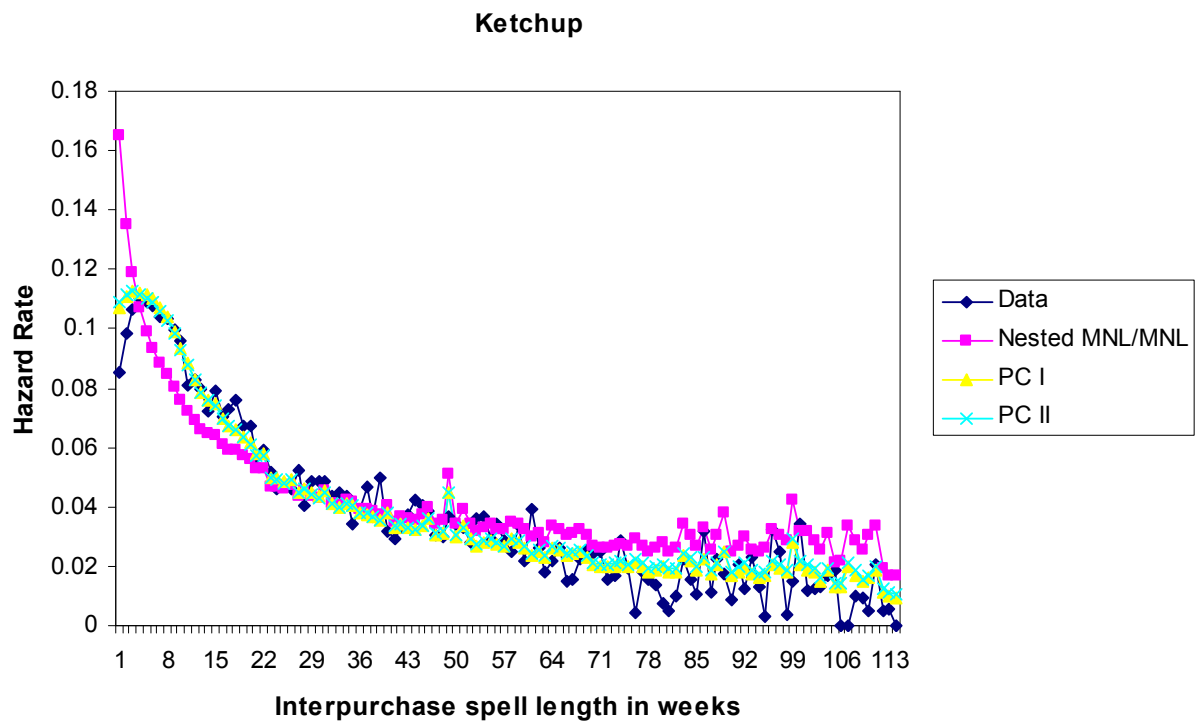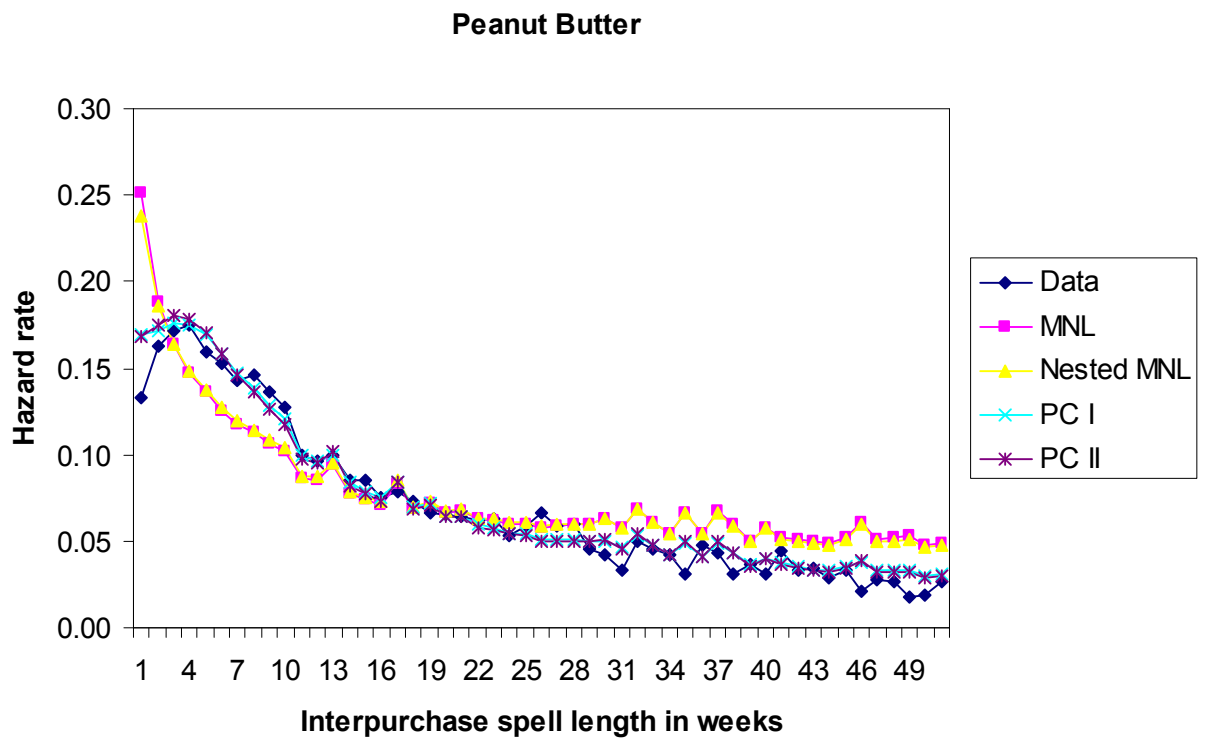
Appendix: Variances and Correlations of the Brand Intercepts

Peanut Butter				
	MNL	NMNL	PC I	PC II
Σ_{11}	2.356	3.275	1.988	2.029
Σ_{22}	2.091	3.013	1.808	1.799
Σ_{33}	2.210	2.934	2.003	1.975
Σ_{44}	1.857	2.595	1.523	1.523
ψ_{21}	-0.054	0.114	-0.023	-0.053
ψ_{31}	0.093	0.219	0.135	0.131
ψ_{32}	0.086	0.233	0.183	0.183
ψ_{41}	0.192	0.369	0.186	0.159
ψ_{42}	0.234	0.413	0.272	0.247
ψ_{43}	0.128	0.300	0.207	0.193

Ketchup				
	MNL	NMNL	PC I	PC II
Σ_{11}	0.811	0.811	0.636	0.664
Σ_{22}	1.696	1.696	1.545	1.571
Σ_{33}	2.179	2.177	2.045	2.057
Σ_{44}	2.420	2.419	2.319	2.243
ψ_{21}	0.200	0.200	0.185	0.204
ψ_{31}	0.070	0.070	0.067	0.077
ψ_{32}	0.692	0.694	0.773	0.766
ψ_{41}	0.032	0.031	0.038	0.074
ψ_{42}	0.390	0.388	0.561	0.536
ψ_{43}	0.443	0.440	0.525	0.509

Σ_{jj} is the variance of α_{ij}. The ψ_{ij}'s are the corresponding correlation matrix elements.

References

Berry, S., Levinshon, J. and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Erdem, T., Imai, S., Keane, M. (2003), "Brand and Quantity Choice Dynamics Under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5-64.

Erdem, T., Keane, M and B. Sun (1999), "Missing Price and Coupon Availability Data in Scanner Panels: Correcting for the Self-Selection Bias in the Choice Model Parameters," *Journal of Econometrics,* 89:1/2, 177-196.

Erdem, T., Keane, M and B. Sun (2006), "A Dynamic Model of Brand Choice When Price and Advertising Signal Product Quality," Working paper.

Fok, D., Paap, R., and Franses, P.H. (2002), "Modeling Dynamic Effects of Promotion on Interpurchase Times," Econometric Institute Report EI 2002/37.

Guadagni, P.M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-238.

Harris, K. and M. Keane (1999). "A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data," *Journal of Econometrics*, 89:1/2, 131-157.

Heckman, James (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Economic Applications*, eds. C. Manski and D. McFadden, Cambridge, MA: MIT.

Helsen, K. and D. C. Schmittlein (1993), "Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models," *Marketing Science*, 12, pp.395-414.

Hendel, Igal, and Nevo, Aviv (2003), "Measuring the Implications of Sales and Consumer Stockpiling," working paper, Northwestern University.

Keane, Michael (1993), "Simulation Estimation for Panel Data Models with Limited Dependent Variables," in *Handbook of Statistics*, G.S.Maddala, C.R. Rao, and H.D. Vinod, eds. New York: Elsevier Science Publishers.

Keane, Michael (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62 (January), 95-116.

Keane, Michael (1997a), "Current Issues in Discrete Choice Modeling," *Marketing Letters* 8:3: 307-322.

Keane, Michael (1997b), "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business and Economic Statistics* 15:3: 310-326.

Maddala, G.S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

McFadden, Daniel (1978), "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, ed. A. Karlquist. Amsterdam, North Holland, 75-96.

McFadden, Daniel (1989), "A Method of Simulated Moments for Estimation of Discrete Choice Response Models Without Numerical Integration," *Econometrica*, 57 (5), 995-1026.

Pakes, Ariel (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54 (4), 755-84.

Sun, B., Neslin, S., Srinivasan, K. (2003), "Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward-looking," *Journal of Marketing Research*, Vol. 40, 389-405.

Sun, Baohong (2005), Promotion Effect on Endogenous Consumption, *Marketing Science*, 3 (24), 430-443.

Wooldridge, Jeff (2003a), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

Wooldridge, Jeff (2003b), "Simple Solutions to the Initial Conditions Problem is Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," Working Paper, Michigan State University.