Title
Consumer learning and evolution of consumer brand preferences

Journal
Quantitative Marketing and Economics, 13(3)

ISSN
1570-7156

Authors
Che, Hai
Erdem, Tülin
Öncü, T Sabri

Publication Date
2015-09-01

DOI
10.1007/s11129-015-9158-x

Peer reviewed

Consumer learning and evolution of consumer brand preferences

Article *in* Quantitative Marketing and Economics · July 2015

DOI: 10.1007/s11129-015-9158-x

3 authors:

Hai Che
University of California, Riverside

Tülin Erdem
New York University

T. Sabri Öncü
Centre for Advanced Financial Research and Learning

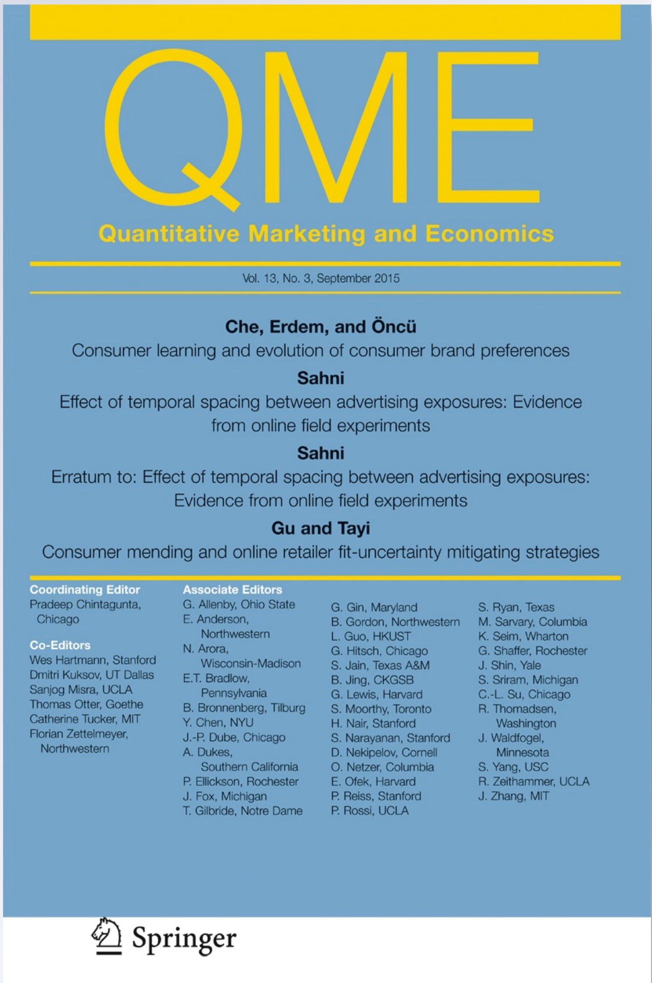
Consumer learning and evolution of consumer brand preferences

Hai Che, Tülin Erdem & T. Sabri Öncü

Quantitative Marketing and Economics
QME

ISSN 1570-7156
Volume 13
Number 3

Quant Mark Econ (2015) 13:173-202
DOI 10.1007/s11129-015-9158-x



Springer

Quant Mark Econ (2015) 13:173–202
DOI 10.1007/s11129-015-9158-x

Consumer learning and evolution of consumer brand preferences

Hai Che[1] · Tülin Erdem[2] · T. Sabri Öncü[3]

Received: 15 April 2013 / Accepted: 15 June 2015 / Published online: 10 July 2015

Abstract We develop a structural dynamic demand model that examines how brand preferences evolve when consumers are uncertain about product quality and their needs change periodically. We allow for strategic sampling behavior of consumers under quality uncertainty and allow for strategic sampling to increase periodically as consumers' needs change periodically. We differ from previous work on forward-looking consumer Bayesian learning by allowing for 1) spill-over learning effects across different versions of products or products in different product categories that share a brand name and 2) duration-dependence in utility for a specific version of a product or product class to capture systematic periodic changes in consumer utility and migration of consumers across product versions or classes. We also assess the evolution of price elasticities in markets where there is consumer quality uncertainty that diminishes over time as consumers get more experienced. We estimate our model using scanner data for the disposable diapers category and discuss the consumer behavior and managerial implications of our estimation and policy simulation results.

Keywords Strategic sampling · Spill-over effects · Duration dependence · Consumer choice under uncertainty · Bayesian learning

Authors are listed alphabetically and contribute equally to the paper.

Electronic supplementary material The online version of this article (doi:10.1007/s11129-015-9158-x) contains supplementary material, which is available to authorized users.

✉ Hai Che
haiche@indiana.edu

Tülin Erdem
terdem@stern.nyu.edu

[1] Kelley School of Business, Indiana University in Bloomington, Bloomington, IN, USA

[2] Stern School of Business, New York University, New York, NY, USA

[3] United Nations Conference on Trade and Development, New York, NY, USA

1 Introduction and background

Previous literature on forward-looking Bayesian learning models has shown that in frequently purchased product categories, consumers may sample brands strategically; that is, they may forgo current utility to get information about brand quality and maximize expected utility over the planning horizon (e.g., Erdem and Keane 1996; Ackerberg 2003; Crawford and Shum 2005; Sun 2005; Hartmann 2006; Osborne 2011). Such dynamic structural demand models predict that in product categories where there is no variety-seeking, consumers switch across brands relatively early on (due to strategic sampling) and later they settle on a small subset of brands once uncertainty is mostly resolved, implying that one observes more switching early on and less switching in later periods, *ceteris paribus*. Thus, standard forward-looking Bayesian learning models suggest that strategic sampling does diminish monotonically over time.

However, even if overall uncertainty diminishes and overall strategic sampling decreases over time, the motivation for sampling for information purposes may increase periodically due to the introduction of new products or the changing needs of consumers. In almost all frequently-purchased consumer packaged goods categories, brands are often re-launched, which may increase strategic sampling. Consumers could expect that the quality of the relaunched brand is highly correlated with the quality of the pre-launch brand but not necessarily perfectly so, which may trigger strategic sampling.

Of particular interest to us are the changing consumer needs that may prompt increased strategic sampling periodically. Consumers may need to migrate from one version of a product to another version of the product or from one product class to another one. In the disposable diapers category, for example, consumers need to switch diaper sizes as their babies grow older. Similarly, consumers switch from formula to baby food when their children grow. Contact lens buyers' prescriptions may change, requiring a switch from one type of lens to another.[1] What is common among these examples is that when consumers migrate from one version of the product (or product class or category) to another version (other product class or category) that share a brand name, consumer quality expectations associated with the new version or the product in the new product class will be only imperfectly correlated with their quality perceptions associated with the product they have prior experience with. Thus, prior quality information obtained through consumer use experience will be only partially applicable to the new version of the same brand or the same brand in the new product category. Thus, when a parent switches to a new size of a disposable diaper brand, or from baby formula to baby food, or when a contact lens user's new prescription requires a different type of lens and the consumer switches contact lens type, consumers again could expect that the quality is highly correlated with other disposable

[1] Another example of prior use experience being only partially relevant when consumer migrate to different versions of products is the camera category, where some consumers switch to more advanced cameras over time and where product usage requires consumer skills (learning-by-doing). Huang (2015) measures the returns to experience on a sample of users of digital cameras, via a measure of their picture quality.

diaper sizes of the same brand, with the other baby food products of the same brand, or with the different types of contact lenses offered by the same contact lens manufacturer, but not necessarily perfectly so. Under these circumstances, the need to switch to new sizes/versions/types of brands (or product categories for an umbrella brand) may trigger periodic strategic sampling.

Thus, our model sheds light on state dependence effects when consumers migrate across product classes and the phenomenon that brand experiences persist across phases of the consumer lifecycle. We should also note that various forms of state dependence can even persist across generations. For example, Anderson et al. (2013) find that there is a strong correlation in the brand of automobile chosen by parents and their adult children using data from the Panel Study of Income Dynamics. The descriptive models they estimate show that this correlation at least partly stems from transmission of brand preferences across generations through state dependence.

Furthermore, in markets where there is quality uncertainty, price elasticities may change over time too which has implications for brand switching over time. If price elastcities increase with reduced quality uncertainty, diminished brand switching due to reduced strategic sampling as consumers gain more experience with brands would be dampened by increased switching due to increased price elastcities in later periods. Thus, overall switching may not decline or even increase when consumer uncertainty is reduced through learning and in later periods consumers become more price elastic. Indeed, when we inspect the brand switching patterns in the scanner panel purchase data for disposable diapers, we do not observe a declining trend in brand switching.[2] To capture both consumer learning and changing price elasticities as consumers become more experienced, it is important to study the behavior of new consumers to a market and observe their behavior over time.

The goal of our research is to model how consumers make decisions in product categories where 1) consumers have quality uncertainty, 2) there are periodically changing needs, prompting migration across versions of products or product classes (e.g., the need to switch to a larger size when the baby grows out of a size in the disposable diapers category); 3) consumer price elasticities may change over time as overall uncertainty diminishes. [3]

To accomplish this goal, we propose and estimate a dynamic structural model of demand using scanner data for the disposable diaper category where one can observe the behavior of new consumers (first time parents) over time. The model allows use experience with a brand's particular size to provide noisy information about another size of the same brand as well. We study use experience spillover effects and the degree to which information from past use experience is retained when consumers migrate across versions of products.

We also study whether price elasticities change over time with diminished uncertainty as babies grow older and parents know more about the brands. In addition, we

[2] In Section 3 (Data and Identification), we talk about these and other descriptive statistics in detail.

[3] Our application is in the disposable diapers category. Heilman et al. (2000) modeled state dependence in disposable diapers category and their descriptive model results indicated that price sensitivities themselves change as a function of use experience in disposable diapers category. To allow for this possibility, we will also allow diaper-size specific price sensitivities in our application.

allow for price sensitivities to be different for smaller versus larger sizes as in categories like disposable diapers or baby food, consumer marginal utility itself may change when they know more about the product category and their own expenses related to their children.

Our modeling approach would be applicable to any frequently purchased product category where consumer needs change periodically or new product introductions and relaunches are frequent. In the case of brand relaunches, an approach similar to the one adopted in this paper can be used to model learning spillover effects between old version of the brand and the re-launched brand. In the case of periodic change in needs, one can think of a number of life-cycle products where our approach would be applicable.[4][5]

Our application is in the disposable diapers category since it is an ideal category for our investigation: it provides an ideal context for learning in general and changing needs and price elasticities over time in particular. First, in the diapers category, potential uncertainty about quality is high because many parents are new to the market and may switch brands to learn about them. Second, unlike in categories such as coffee or detergent where consumers may be using the category for many years and have very well-established tastes, in the disposable diapers category price elasticities (and sensitivities) may evolve as well. Thus, by focusing on first-time parents,[6] one can observe the evolution of consumer preferences and choices as the parents get more experienced.

We should also note that an expanded version of our model can be applied to high-tech durables too where new generation of products are routinely introduced and where learning spill-over effects are to be expected across the old generation and new generation of products (Erdem et al. 2005; Goettler and Gordon 2011; Gowrisankaran and Rysman 2012). The set-up in this paper is different since needs change exogenously in our set-up (babies grow, people get older) and the time to switch to the next size or type of product is exogenous whereas in the high-tech durables case the time to adopt or upgrade is endogenous as well. Thus, our set-up would need to be expanded to allow for endogenous timing decisions to be applicable to such markets.

The forward-looking structural demand model proposed and estimated in this paper accounts for changing needs and price elasticities/sensitivities over time and the possibility that motivation for strategic sampling may increase periodically as the needs change. To our best knowledge, this is the first dynamic structural demand model with forward-looking consumers with experience spill-over effects. Our results show that 1)

[4] For example, Li et al. (2005) investigate customer purchase patterns for products that are marketed by a large bank. To do so, they estimate a multivariate probit model to investigate how customer demand for multiple products evolves over time and its implications for the sequential acquisition patterns of naturally ordered products (e.g., open a credit card account first when young and then applying for a mortgage). They do not model learning explicitly but our modeling approach can be adopted and adapted to model learning in such settings as well.

[5] We thank to an anonymous reviewer for the insight that our framework is equally applicable to any repeated product choices from the same (umbrella) brand over the customer's life-cycle.

[6] Focusing on first-time parents also alleviates greatly the initial conditions problem that all dynamic models are subject to.

consumer experience of a particular size of a particular brand serves not only as a quality signal for that size but also for other sizes of the same brand; 2) consumer brand-size preference is duration dependent in such a way that it first increases and then decreases with the time that consumers stay with a particular brand-size; and 3) consumer price elasticities (as well as sensitivities) change when their babies grow older. Finally, we conduct policy experiments to describe how marketers may tailor their marketing activities when consumer needs change periodically.

2 The model

2.1 Overview

We model household behavior in a market in which households may be uncertain about product quality and risk-averse. We allow households to use their use experience as a signal of product quality; that is, they learn about product quality through use experience and update their expectations in a Bayesian manner. However, we go beyond the usual Bayesian learning models in that we account for spill-over effects of the signals from one subset of the product category to another by allowing correlations between signals from different subsets of the category.[7] In our study of purchases in the diapers market, we accomplish this by projecting the information provided by use experience in a size of a diaper brand onto other sizes of the same brand. We allow households to be forward-looking in the sense that they maximize the expected value of the aggregate present values of their future utilities over a planning horizon. This leads to strategic-sampling as in Erdem and Keane (1996). However, we go beyond the now standard strategic-sampling model of Erdem and Keane (1996) in that we account for the fact that in certain markets consumer needs may change periodically. In the specific case of diapers category, for example, there is the need to switch to a bigger size as the baby grows older, which may prompt temporary increased strategic sampling. This is so because use experience signals (quality information through use experience) associated with a smaller size of a brand may not be perfectly correlated with use experience signals (quality information through use experience) associated with a larger size. Finally, we also allow for price sensitivities to be different for smaller versus larger sizes to capture the possibility that parents' price sensitivities may change as they get more experienced with the product category. A formal description of what we propose is below.

[7] Previous papers that incorporated spill-over learning effects across products or products attributes assumed myopic agents and did not model duration dependence in utility (e.g., Erdem 1998; Coscelli and Shum 2004 and Chan et al. 2013). One exception is Dickstein (2011), who, like us, allows for forward-looking behavior. Dickstein considers a model with forward-looking physicians facing a multi-armed bandit problem, where a physician is uncertain about his patients' intrinsic preference for drugs' characteristics, and he makes use of patients' total utility of consuming a drug in time t to update his belief about their preferences. The proposed model does not allow for risk-averse behavior or evolving needs.

2.2 Consumer utility

Consider a set of households H={$h|h$=1,…,H} purchasing from a product category. Let T={$t|t$=1,…,T} be the time period (week), J={$j|j$=1,…,J} be the set of brands available in the category, and K={$k|k$=1,…,K} be the set of sizes of all of the brands.[8] Since we will estimate the model using data from disposable diapers category, we let k stand for size but k could label also different versions of a product that shares an umbrella brand (different types of contact lenses) or different products in different categories or sub-categories that share an umbrella brand (e.g., Gerber formula, Gerber baby food, Gerber snacks for older kids). The choice set for each household is specified such that household h's choice set does not include all the K sizes since sizes too large or too small would not meet babies' needs. Specifically, at each t when/if a purchase occurs, we assume that the choice set includes the size purchased in the previous purchase occasion,[9] as well as the sizes adjacent to it. For example, if size 1 is the size purchased in the previous purchase occasion, then choice set consists of the sizes 1 and 2 or if size K is the size purchased in the previous purchase occasion, then the choice set consists of the sizes K-1 and K, whereas for any interior size (size 2,3, or 4 in our dataset) the choice set consist of the size purchased in the previous purchase occasion and its two neighbors such as 1, 2 and 3 if size 2 is the size purchased in the previous purchase occasion. Let us ignore this for a while and start with some auxiliary utility functions of the form[10]

$$\hat{U}_{htjk} = \alpha_{hm_k} p_{htjk} + w_h \left(Q_{Ehtjk} - r_h Q^2_{Ehtjk} \right) + \mu_h \left(D_{h,t-1,k} - s_k D^2_{h,t-1,k} \right) + \varepsilon_{htjk} \quad (1)$$

where brand j= $1,...,J$, size k=1,2,…,K, household h= 1,...,H, and t= 1,…T. In the above, p_{htjk} is the price faced whereas Q_{Ehtjk} is the quality experienced by household h of size k of brand j at time t. The parameters w_h and r_h are the respective measures of the utility weight of and degree of risk aversion to the unobservable product quality, both expected to be positive. The parameter α_{hm_k} is the price coefficient, expected to be negative. Here the subscript m_k captures diaper size. We allow for size specific price coefficients and allow these coefficients to be different between small sizes (which are bought by parents with little experience with the category) and large sizes (which are bought by parents with more experience with the category). More specifically, m_k will denote a small size when k= 1, 2 or 3, say, m_k =1; and m_k will denote a large size when k≥4, say, m_k =2; and we note that in our dataset there are five sizes, that is, in our dataset K=5.

Allowing for price sensitivities to be small versus large size-specific enables us to capture whether and how these sensitivities evolve over time as babies grow. We should note that price elasticities may change over time in markets with quality uncertainty as consumers become more experienced with the category and the qualities of individual brands. This does not mean that price coefficients (price sensitivities) would change.

---

[8] Here "size" refers to the size of the individual diaper, such as the newborn size, and there are 5 sizes. It does not refer to the package size.

[9] When a household does not make a purchase during week t, we assume his choice set included the *last* purchased size and its adjacent sizes as discussed in the above text.

[10] Please note that we tried data on features as a control variable (we do not have display variable). However, the feature variable was statistically insignificant in the structural models we estimated so we did not use it as a control.

On the other hand, price sensitivities themselves may change in those categories where consumers are new to a whole product category and entry into the product category corresponds to a specific life-stage of the consumer. For example, in baby foods, toys or disposable diapers as first-time parents get a better sense about their expenses in these categories or as their marginal utility of income changes as their overall expenses associated with their children change.[11] Since our application will be in the diapers category, we allow for this possibility. Finally, ε_{htjk} is the taste shock that becomes known to the household at time t but is unknown to the econometrician. We specify the distributional properties of ε_{htjk} later in the text.

Lastly, we note that the term $\mu_h(D_{h,t-1,k}-s_k D^2_{h,t-1,k})$ in the Eq. (1) captures duration dependence in a size, where D_{htk} is the number of periods in the k^{th} size until time t since the household h made the first purchase of the k^{th} size. We call D_{htk} the size duration variable. This term captures changing needs of consumers over time. In the specific case of diapers, we expect that the time spent in a particular size would increase the expected utility initially but that as the baby gets closer to growing out of the current size, the positive influence of the time spent in the current size would diminish gradually[12] so this specification allows for this possibility. While the duration weight μ_h is heterogeneous across households, the size specific location coefficients s_k are assumed to be the same for all of the households. Provided that the duration weight μ_h is positive in the mean and that $0<s_k<1$, consumer utility increases with the size duration for as long as $D_{htk}<1/\,s_k$ and, thereafter, it starts to decrease.[13]

Let us now get back to the choice set issue we mentioned above and denote by K_{ht} the size purchase of household h on purchase occasion t. To address the choice set issue – with the help of our above auxiliary utility functions – we specify the individual utility functions as follows:

$$U_{hjkt} = \begin{cases} \hat{U}_{hjkt}, & \text{if } K_{h,t-1}-1\leq k\leq K_{h,t-1}+1 \\ -\infty, & \text{otherwise} \end{cases} \tag{2}$$

This utility specification ensures that at each t when/if a purchase occurs, the choice set includes only the size purchased in the previous purchase occasion and the sizes

[11] Indeed, Heilman et al. (2000) found that price sensitivities are time-varying and a function of cumulative use experience in the diapers category. We should also note that descriptive models (such as varying parameter models) have shown evidence of changing price sensitivities in few other frequently purchased product categories as well (e.g., Mela et al. 1997).

[12] When babies first grow into a size, the fit to the new size may be not perfect, so utility may first increase as time passes and the fit gets better. Then, when the baby is about to grow out of a size, the fit may again diminish. The duration dependence term is set up so that after a consumer has been in size k for a while, she is more likely to move to k+1 than to k-1. We thank an anonymous reviewer for this insight.

[13] We also tried a utility specification where Equation 1 has a last brand purchase dummy to capture any one-lag state dependence effects not related to learning (e.g., switching costs (Dubé et al. 2009) or preference inertia (Che et al. 2007; Shin et al. 2012)). Indeed, Osborne (2011) found that in frequently purchased product categories there are both learning and switching costs. The coefficient of the last purchase dummy is identified in such contexts as discussed on Osborne (2011). We found that lagged purchase dummy has statistically significant but (size-wise) very small effect, and the results were very similar between the two models. We turned that component off for three reasons. First, we turned it off for parsimony since our model has already quite a few "moving-parts" as it focuses on evolution of needs (the need to switch to a different size), learning across sizes and changing price sensitivities. Second, a learning model fits our data better than a model with no learning but a lagged dependent variable or with a weighted average of past purchases variable. Third and most importantly, these lagged purchase variables added to learning models are behaviorally difficult to interpret.

adjacent to it, as well as the no-purchase option, of course. If there is no purchase at any given time, we set K_{ht}=$K_{h,t-1}$ and proceed in the above manner to the next purchase occasion. The significance of this utility specification is that the growth of babies is exogenous to the model so that size changes occur exogenously. Although the guardian of the baby observes the growth of the baby, this process is not observable by the econometrician. The econometrician observes only the previously made purchases. Since it would be unrealistic to expect a baby to grow two or more sizes between two purchase occasions and it is possible that an early attempt of a larger size could have been made to experiment with the size on any occasion, we restrict the choice set as above.

2.3 Consumer expected utility

Let Q_{jk} denote the unobserved *true* quality of size k of brand j about which the *experienced* quality Q_{Ehtjk} fluctuates. Fluctuations of the experienced quality Q_{Ehtjk} around the unobserved true quality Q_{jk} may occur for many reasons. One possibility is the variability of product quality across batches of products, to name one. Another possibility is the context dependence of consumer use experiences, a more plausible explanation for product categories covered by scanner data typically, to name another. And there may be many other reasons. Irrespective of the reasons, however, we formulate the fluctuations in use experience by assuming that each use experience provides a noisy but unbiased signal of quality according to Q_{Ehtjk}= Q_{jk}+ ξ_{htjk}, here ξ_{htjk}~$N(0,\sigma_\xi^2)$ and σ_ξ^2 is the *experience variability* (the reciprocal of which denotes the precision of the use experience information or the quality signal associated with the use experience). We assume further that ξ_{htjk} are identically and independently distributed.

It is evident that the quality *experienced* by household h of size k of brand j at time t, that is, Q_{Ehtjk}, does not need be the same as the quality *perceived* by the same household of the same size of the same brand at the same time, that is, Q_{htjk}. We assume that the *perceived* quality is given by Q_{htjk}= Q_{jk}+υ_{htjk}, where the perception errors are distributed as υ_{htjk}~$N\left(0,\sigma^2_{\upsilon_{htjk}}\right)$. The perception variance (variance of quality beliefs) in period t, $\sigma^2_{\upsilon_{htjk}}$, is updated after a purchase is made in period $t-1$.

We also allow initial variance of quality beliefs $\sigma^2_{\upsilon_{h0jk}}$ to be a function of purchases in the category during the initialization period (in our sample, 27 weeks) as a crude measure of heterogeneity in initial uncertainty across households. Thus, we assume that

$$\sigma^2_{\upsilon_{h0jk}} = \lambda_0 + \lambda_1 \times N_h + \zeta_{h0jk}$$

where N_h is the number of purchases in the initialization period of the household h.

Finally, since we have assumed that the use experience signals are unbiased, we have Q_{htjk}= $E[Q_{Ehtjk}|I_{h,t-1}]$=$E[Q_{jk}|I_{h,t-1}]$.

Prior to making a purchase decision in period t, household h forms the $I_{h,t-1}$ -conditional expectations of the auxiliary utility functions $\hat{U}_{htjk}$ for each of the sizes of all brands as follows:

$$\begin{aligned} E\left[\hat{U}_{htjk}|I_{h,t-1}\right] = {} & \alpha_{hm_k}p_{htjk} + w_h Q_{Ehtjk} - w_h r_h Q^2_{Ehtjk} - w_h r_h\left(\sigma^2_{\upsilon_{htjk}} + \sigma^2_\xi\right) \\ & + \mu_h\left(D_{h,t-1,k} - s_k D^2_{h,t-1,k}\right) + \varepsilon_{htjk} \end{aligned} \quad (3)$$

Since $Q_{htjk} = Q_{jk} + \upsilon_{htjk}$, the $I_{h,t-1}$ -conditional expected auxiliary utilities $E[\hat{U}_{htjk}|I_{h,t-1}]$ depend not only on the unobservable product qualities Q_{jk} but also on the perception errors υ_{hjkt}. Rewritten explicitly:

$$E\left[\hat{U}_{htjk}\middle|I_{h,t-1}\right] = \alpha_{hm_k} p_{htjk} + w_h\left(Q_{jk} + \upsilon_{htjk}\right) - w_h r_h\left(Q_{jk} + \upsilon_{htjk}\right)^2 - w_h r_h\left(\sigma^2_{\upsilon_{htjk}} + \sigma^2_{\xi}\right) + \mu_h\left(D_{h,t-1,k} - s_k D^2_{h,t-1,k}\right) + \varepsilon_{htjk} \tag{4}$$

As is evident from Eq. (4), there are two sources of consumer uncertainty: the first is the perception variability $\sigma^2_{\upsilon_{htjk}}$ (the variance of consumer quality beliefs) whereas the second is the experience variability σ^2_{ξ}. Although the perception variability diminishes with use experience in our model, the experience variability does not.

Additionally, we specify the utility of no purchase as $U_{ht00} = \gamma_0 + \gamma_h INV_{ht} + \varepsilon_{ht00}$, where INV_{ht} is the household's inventory of the product category. We model inventory as $INV_{ht} = INV_{ht-1} + q_{ht-1} - C_h$, where q_{ht-1} denotes the quantity of category purchased by household h at purchase date t-1 and C_h denotes household h's consumption rate (Bucklin and Gupta 1992). We measure the consumption rate, C_h, as the average weekly consumption of diaper and it is computed as the total number of pieces of diaper purchased by household h divided by the number of weeks in the sample period. Lastly, γ_h is the household-specific inventory weight and γ_0 is the intercept of the no-purchase utility.[14] [15]

Finally, we let α_h, w_h, and r_h be heterogeneous across consumers and, following Heckman (1981), adopt a latent class approach. (α_m, w_m, r_m) as well as the associated population type proportions π_m for each of the consumer segments $m = 1,2,\ldots,M$. Many papers that involve Dynamic Programming models utilize a latent class approach to capture unobserved heterogeneity[16] since imposing a continuous distribution for heterogeneity would imply solving the Dynamic Programming problem for each household which is difficult to accomplish. We acknowledge that the latent segments themselves may not be completely homogenous, and if so, then since heterogeneity will be underestimated, the learning parameters would be biased (see Shin et al. 2012).

[14] The estimate of γ_0 was not statistically significant so we did not report them in the result tables.

[15] We should also note that we model brand choice and purchase incidence but do not model quantity choice (rather we model the impact of inventories on the probability of purchase incidence in a descriptive way). Previous papers on forward-looking dynamic structural models focused either on quality expectations, learning and strategic sampling in the context of brand choice or on both brand and quantity choice and price expectations but assumed away quality learning and strategic sampling since it is not feasible to model both processes in one structural model that explicitly allows for both quality and price expectations (Erdem et al. (2008). Furthermore, Ching et al. (2014) do so in a semi-structural model and find that in the diapers category the quality learning effects are significant whereas the price expectation effects are not for first-time parents.

[16] Hartmann (2006) and (2010) are two exceptions that allow for richer unobserved heterogeneity structures.

2.4 Consumer learning

We assume for each household $h \in \mathrm{H}$, brand $j \in \mathrm{J}$, and size $k \in \mathrm{K}$ that the initial perception errors υ_{h0jk} are correlated across sizes and their correlation matrix is given by:

$$R = \begin{bmatrix} 1 & \rho_{12} & 0 & . & . & . & . & 0 \\ . & 1 & \rho_{23} & . & . & . & . & 0 \\ . & . & 1 & \rho_{34} & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & 1 & \rho_{K-2,K-1} & 0 \\ . & . & . & . & . & . & 1 & \rho_{K-1,K} \\ . & . & . & . & . & . & . & 1 \end{bmatrix}$$

where $\rho_{h0kl}=\rho_{kl}$ are the initial correlations across perception errors between sizes k and l. These capture correlations in consumer perceptions across sizes. This specification indicates that only the adjacent sizes are correlated and that the initial size correlations are uniform across brands and households, by assumption.[17] We denote by ρ_{htkl} the time $t \geq 1$ correlation coefficients between the sizes of the brands, which are updated over time as we describe below.

Since υ_{htjk} are correlated across sizes, we have the following relationships across different sizes of the brands for any $t \geq 1$:

$$Q_{Ehtjk} = \kappa_{htjk|l} Q_{Ehtjl} + \eta_{htjk|l} 1_{\{l \neq k\}}, \quad \eta_{htjk|l} \sim N\left(0, \sigma^2_{\eta_{htjk|l}}\right) \tag{5}$$

In mathematical terms, this is the linear projection of one vector on another, and κ can be thought of the ordinary least square coefficient for Q_{Ehtjl}. Therefore,[18]

$$\kappa_{htjk|l} = \frac{\sigma_{\upsilon_{h,t-1,jk}} \sigma_{\upsilon_{h,t-1,jl}} \rho_{h,t-1,kl}}{\sigma^2_{\upsilon_{h,t-1,jl}} + \sigma^2_{\xi} 1_{\{l \neq k\}}} \tag{6}$$

and

$$\sigma^2_{\eta_{htjk|l}} = \frac{\left(\sigma^2_{\upsilon_{h,t-1,jk}} + \sigma^2_{\xi} 1_{\{l \neq k\}}\right)\left(\sigma^2_{\upsilon_{h,t-1,jl}} + \sigma^2_{\xi} 1_{\{l \neq k\}}\right) - \sigma^2_{\upsilon_{h,t-1,jk}} \sigma^2_{\upsilon_{h,t-1,jl}} \rho^2_{h,t-1,kl}}{\sigma^2_{\upsilon_{h,t-1,jl}} + \sigma_{\xi}{}^2 1_{\{l \neq k\}}} \tag{7}$$

whereas $1_{\{A\}}$ is the indicator function that returns 1 if the statement A is true and 0 otherwise.

To interpret the above formulae, let us consider the zero correlation case (ρ=0) first. In this case, the use of a size of a particular brand provides no information on the quality of other sizes of the same brand. In other words, the information provided by this use signal is pure noise and, hence, consumers do not update their quality

[17] In our sample, consumers only buy products across adjacent sizes and do not purchase more than one size up or down. The correlation matrix between adjacent sizes is set up to reflect the fact that there is a natural sequence in consumer's purchase of different sizes. In a more general setting, to model the correlation between product classes (e.g., sedan, SUV and truck), it would be useful to generalize the correlation matrix to allow more flexible spill-over effects. We thank an anonymous reviewer for this insight.

[18] More details of the learning model are provided in a Technical Appendix available upon request from the authors.

perceptions of the other sizes of the brand. Only the quality perception of this particular size of the brand gets updated. This is the first extreme.

The second extreme is when the correlation is perfect (ρ=1). In this case, consumers do not distinguish between different sizes of the brands. That is, they treat all sizes of any specific brand to be of equal quality. In other words, the quality perceptions of all sizes of each brand are the same and get updated for the brand – not for sizes of the brand – in the same manner no matter what size of the brand is used.

The reality should be somewhere between these two extremes and this is where our formulation comes into play. It is clear from Eqs. (5) to (6) that as long as ρ_{htkl} are non-zero, use experience in a particular size of a brand provides information for other sizes of the same brand as well. Furthermore it is clear from Eq. (6) that $0<|\kappa_{htjk|l}|<1$ when $k\neq l$ whereas $\kappa_{htjk|k}=1$. This implies that the noisy information provided by the use experience in a different size of a brand is less than the information provided by the use experience in the current size, as it should be, unless consumers view all sizes of the brand to be of equal quality.

It is evident that our approach can be employed to study the spill-over effects of signals from one domain to another in the context of forward-looking consumers. As previously indicated, all previous learning models with spill-over effects have assumed myopic consumers, with the exception of a working paper by Dickstein (2011), whose method requires risk-neutrality (and Dickstein's model does not allow for changing needs and price sensitivities). Our model allows for risk-aversion as well, which has been shown to hold empirically in previous learning papers.

Let us now denote by J_{ht} brand choice – and recall that K_{ht} is the size choice – of household h on purchase occasion t. With this definition and the above relationships between the sizes k and l, we are now ready to describe how the consumers update their perceived qualities in our model. We assume that after buying size l of brand j on purchase occasion t, household h updates the priors about the mean quality of size k of brand j using Bayesian updating rules (see, for example, DeGroot 1970). Therefore, we have the following updating equations for $t>1$:

$$\upsilon_{htjk} = \kappa_{htjk|K_{h,t-1}}\upsilon_{h,t-1,jk} + \frac{\kappa^2_{htjk|K_{h,t-1}}\sigma^2_{\upsilon_{h,t-1,jk}}}{\kappa^2_{htjk|K_{h,t-1}}\left(\sigma^2_{\upsilon_{h,t-1,jk}}1_{\{J_{h,t-1}\neq j\}}+\sigma^2_{\xi}\right)+\sigma^2_{\eta_{htjk|K_{h,t-1}}}1_{\{K_{h,t-1}\neq k\}}} \cdot\left\{\kappa_{htjk|K_{h,t-1}}\left(\xi_{htjk}-\upsilon_{h,t-1,jk}\right)+\eta_{h,t-1,jk|K_{h,t-1}}1_{\{K_{h,t-1}\neq k\}}\right\}1_{\{J_{h,t-1}\neq j\}} \tag{8}$$

$$\sigma^2_{\upsilon_{htjk}} = \frac{\kappa^2_{htjk|K_{h,t-1}}\left(\kappa^2_{htjk|K_{h,t-1}}\sigma^2_{\xi}+\sigma^2_{\eta_{htjk|K_{h,t-1}}}1_{\{K_{h,t-1}\neq k\}}\right)\sigma^2_{\upsilon_{h,t-1,jk}}}{\kappa^2_{htjk|K_{h,t-1}}\left(\sigma^2_{\upsilon_{h,t-1,jk}}1_{\{J_{h,t-1}\neq j\}}+\sigma^2_{\xi}\right)+\sigma^2_{\eta_{htjk|K_{h,t-1}}}1_{\{K_{h,t-1}\neq k\}}} \tag{9}$$

and

$$\rho_{hjklt} = \begin{cases} \rho_{hjkl,t-1} & J_t \neq j \\ \dfrac{\mathrm{cov}\left(\upsilon_{hjkt}, \upsilon_{hjlt}\right)}{\sigma_{\upsilon_{hjkt}} \sigma_{\upsilon_{hjlt}}} & J_t = j \end{cases} \tag{10}$$

where,

$$\begin{aligned} \mathrm{cov}\left(\upsilon_{hjkt}, \upsilon_{hjlt}\right) &= \left(1-\delta_{htjk|K_{h,t-1}}\right)\left(1-\delta_{htjl|K_{h,t-1}}\right)\kappa_{htjk|K_{h,t-1}}\kappa_{htjl|K_{h,t-1}}\sigma^2_{\upsilon_{h,t-1,jk}} \\ &\quad + \delta_{htjk|K_{h,t-1}}\delta_{htjl|K_{h,t-1}}\kappa_{htjk|K_{h,t-1}}\kappa_{htjl|K_{h,t-1}}\sigma^2_{\xi}, \quad \text{and} \\ \delta_{htjk|K_{h,t-1}} &= \frac{\kappa^2_{htjk|K_{h,t-1}}\sigma^2_{\upsilon_{h,t-1,jk}}}{\kappa^2_{htjk|K_{h,t-1}}\left(\sigma^2_{\upsilon_{h,t-1,jk}} 1_{\{J_{h,t-1}\neq j\}} + \sigma^2_{\xi}\right) + \sigma^2_{\eta_{htjk|K_{h,t-1}}} 1_{\{K_{h,t-1}\neq k\}}} \end{aligned} \tag{11}$$

We observe from the above that both the variances and correlations are decreasing to zero with the number of purchases, as is evident from Eqs. (8) through (11). Of course, this does not mean that consumers will eventually learn the product qualities of the brands and sizes with certainty since that would require infinitely many purchases of diapers from each brand and size.

2.5 Consumer expected utility maximization over the planning horizon

Recalling that J_{ht} and K_{ht} are the respective brand and size purchases of household h on purchase occasion t, we suppose that the forward-looking household h solves the following dynamic programming problem:

$$\max_{\{(J_{ht}, K_{ht}) \in J \times K \cup (0,0) | t = 1,2,\ldots T_H\}} E\left[\sum_{t=1}^{T_H} \beta^{t-1} E\left[U_{ht\,J_{ht}K_{ht}} \middle| I_{h,t-1}\right] \middle| I_{h0}\right] \tag{12}$$

where β is the one-period discount factor. We choose β=0.995.[19] The planning horizon T_H may go beyond the end of observation period T; that is, we may have $T_H > T$.

To solve the consumer dynamic programming problem (12), we apply Bellman's principle to solve this problem by finding value functions corresponding to each

[19] We fixed the weekly discount factor β at 0.995 since the discount factor is often difficult to identify even when certain variables can be found that affect expected payoffs but nor current utility (that is, exclusion restrictions may exist). For example, Erdem and Keane (1996) found, in a similar but simpler model, the likelihood was quite flat over a range of discount factors in the vicinity of 0.995, which was the case for us too. We estimated the model with few different weekly discount factors but the results were not very sensitive to the exact value of the discount factor. Please note that the best way to identify the discount factor is either to find contexts where proper exclusion restrictions and practical identification exist (e.g., Chung et al. (2014)) or use (experimental or field) data that has information on behavior both in static and dynamic contexts/regimes to pin down the discount factor (e.g., Yao et al. (2012)) but we do not have such data. There are indeed very few cases where such data are available.

alternative choice. The value of choosing alternative j,k, where $j=0,1,..,J$, $k=1,\ldots 5$, at purchase occasion t is:

$$V_{htjk}(I_{ht}) = E\left[U_{htjk}|I_{ht}\right] + \beta\cdot E\left[\max_{l,m} V_{h,t,l,m}\left(I_{h,t+1}\right)|I_{ht}\right] \tag{13}$$

subject to the terminal condition

$$V_{hT_Hjk} = E\left[U_{hT_Hjk}|I_{h,T_H-1}\right] \tag{14}$$

To compute the V_{htjk}, we must compute the above Emax functions appearing in recursion relation (13) for each of the alternatives. This is not an easy task. However, if we assume that the stochastic taste shocks ε_{hjkt} to the expected utilities $E[U_{htjk}|I_{h,t-1}]$ are identically and independently extreme value distributed, then we can obtain closed form expressions for the above $\{I_{h,t-1},J_{ht}=j,K_{ht}=K\}$ conditional Emax functions as detailed in Rust (1987). We assume now that this is the case.

To close this subsection, recall that size changes are exogenous in our model and that the households need to solve the dynamic programming problem (12) without knowing ahead of time when their size needs will change. Therefore, although the size changes occur exogenously, they are not deterministic but random variables. Our utility specification allows us to handle these random size changes in a straightforward way. More specifically, recall that we as econometricians observe the solution of the consumer's dynamic programming problem and estimate (or reverse engineer) his/her utility function. Since we are solving the Bellman's equation on any purchase occasion for that occasion, we know the full information set of the consumer up to that occasion. Once the random future value functions based on the utility specification are generated, the current value functions of the current purchase occasion are computed from the Eq. (13), or alternatively, from the Eq. (14), if the current purchase occasion is the last purchase occasion.

2.6 Consumer choice probabilities and the likelihood function

Since we have assumed that the stochastic taste shocks ε_{hjkt} are identically and independently extreme value distributed, the consumer choice probabilities are the conditional logit probabilities of McFadden (1974). When consumers are myopic, the period t choice probabilities of household h for latent class m conditioned on the period t perception errors vector $\upsilon_{ht}=\{\upsilon_{htjk}|(j,k)\in \mathrm{J}\times\mathrm{K}\cup(0,0)\}$ are:

$$\theta_{htjk}(\Theta_m|\upsilon_{ht}) = \frac{\exp\left\{E\left[U_{htjk}|I_{h,t-1}\right]\right\}}{\sum_{(m,n)\in \mathrm{J}\times\mathrm{K}\cup(0,0)} \exp\left\{E\left[U_{htmn}|I_{h,t-1}\right]\right\}} \tag{15}$$

while when consumers are forward-looking, they are

$$\theta_{htjk}(\Theta_m|\upsilon_{ht}) = \frac{\exp\left\{E\left[U_{htjk} + \beta E\left[V_{h,t+1}|I_{h,t-1}, J_{ht}=j, K_{ht}=k\right]|I_{h,t-1}\right]\right\}}{\sum_{(m,n)\in \mathrm{J}\times\mathrm{K}\cup(0,0)} \exp\left\{E\left[U_{htmn} + \beta E\left[V_{h,t+1}|I_{h,t-1}, J_{ht}=m, K_{ht}=n\right]|I_{h,t-1}\right]\right\}} \tag{16}$$

where $\Theta_m=\{\alpha_m,\beta_m,r_m,\mu_m,\gamma_m,w_m,r_m,\pi_m,\sigma_\xi,\sigma_\upsilon,\sigma_\eta,\varphi,Q,\lambda_0,\lambda_1,\rho,s\}$ is the class m parameter vector in which $Q=\{Q_{jk}|(j,k)\in \mathrm{J}\times\mathrm{K}\}$, $\rho=\{\rho_{kl}|l=k+1;k=1,K-1\}$ and $s=\{s_k|k=1,K\}$.

Irrespective of whether households are myopic or forward-looking, however, the class m $\upsilon_h=\{\upsilon_{ht}|t\in \mathrm{T}\}$ conditional likelihood function of household h associated with the purchases made over the observation period T is:

$$L_{hm}(\Theta_m|\upsilon_{ht}) = \prod_{t=1}^{T} \prod_{(j,k)\in J\times K\cup(0,0)} \theta_{htjk}(\Theta_m|\upsilon_{ht})^{Y_{htjk}} \tag{17}$$

where $Y_{htjk}=1$, if household h bought size k of brand j on purchase occasion t, while $Y_{htjk}=0$ otherwise.

Had the consumer perceptions errors vector υ_h been observable to the econometricians, the above υ_h conditional likelihood function of household h for each of the latent classes $m=1,2,\ldots,M$ would have sufficed. However, because the consumer perceptions errors vector υ_h is not observable to us, we need to work with the unconditional likelihood function for latent class m given by:

$$L_{hm}(\Theta_m) = \int_{\upsilon_h\in\Omega} L_h(\Theta_m|\upsilon_h) f(\upsilon_h) d\upsilon_{h1} d\upsilon_{h2} \ldots d\upsilon_{hN} \tag{18}$$

where $f(\upsilon_h)$ is the joint distribution, N is the length, and Ω is the obvious domain of the household perception errors vector υ_h. Since it is impossible to carry out the above integration analytically, we have to resort to numerical techniques. We use the interpolative regression method developed by Keane and Wolpin (1994) and used in many previous forward-looking learning models, including Erdem and Keane (1996).

Once $L_{hm}(\Theta_m)$ is computed by employing the simulation technique mentioned above, we can then calculate the likelihood function of household h from:

$$L_h(\Theta) = \sum_{m=1}^{M} \pi_m L_{hm}(\Theta_m) \tag{19}$$

where $\Theta=\{(\Theta_m,\pi_m)|m=1,2,\ldots,M\}$ is the overall parameter vector. The likelihood function of the entire sample we maximize to estimate the parameter vector Θ is then:

$$L(\Theta) = \prod_{h=1}^{H} L_h(\Theta) \tag{20}$$

3 Data and identification

We use scanner panel data from a large national grocery chain on household purchases of disposable diapers between May 2005 and May 2007 in one store in the San Francisco Bay area. The store is located in a mountainous area and has no other large grocery competitors (stores from other retailers or from the retailer itself) or grocery supercenters (e.g., Target or Wal-Mart) within a 5-mile radius. One potential problem of using data from one retail chain is that the observations of consumers shopping in competing stores are unavailable. Using data from one store that has no competing stores nearby, however, helps to reduce such possible bias. The data record all store visits for 105 weeks in 2005–07 for a panel of over 50,000 households in Northern California. Both the brand purchased and price paid are recorded.

The disposable diapers category is an ideal category for our purposes since: 1) the potential for strategic sampling is high in this category as we are studying the choices of first-time parents who are identified by household demographics information (e.g., children's ages and the number of children) and 2) this is a category where household needs change periodically due to the need to switch to a bigger size when the babies get older, 3) initial conditions problem that is relevant for all dynamic brand choice models but even more problematic for learning models is less an issue here since the sample of households analyzed are new to this market as described more in detail below.

We analyze three brands (Pampers, Huggies, and the store brand, which together have a 98 % market share) and use the purchase selection procedure (Gupta et al. 1996) to retain households purchasing only these 3 brands in the category. To capture learning behavior over time and minimize initial conditions problem, we then focus on first-time parents who have made at least 22 purchases in 105 weeks. Given that mean purchase frequency is about 3 to 5 weeks (it varies by size), this selection criteria allows us to exclude first-time parents whose child is about to grow out of diapers. Among the resulted 1007 regular diaper-user households who made at least 22 purchases, we identify 365 households who are first time parents. We identify first-time parenthood using the total number of children and number of children in each age bracket information available in the data. We define the first-time parents as parents who have 1) *only one* child; and, 2) the child is under 3. Then, we then use a random number generator to assign these households to the calibration and validation samples. In this way, 191 first-time parent households are selected for calibration and 174 for validation.

The observations in the first 27-week (the initialization period) are used to allow for heterogeneity in consumer initial uncertainty. As indicated in Section 2.2, we allow initial perceived variance to be a function of purchases in the category during the initialization period. As the estimation sample covers 78 weeks, the calibration and holdout samples have 9102 and 7918 observations, respectively. For the 191 households in our estimation sample, the average number of purchases is 42, which is quite similar to that in Heilman et al. (2000), and the standard deviation is 17 purchases. The average lengths of interpurchase time between purchases are 1.91, 2.26, 3.64, 3.89, and 4.15 weeks respectively for sizes 1 through 5.

Table 1 reports descriptive statistics for different brands and sizes for the calibration sample.

Note that Pampers and Huggies each has 47~48 % of the total market share, and the store brand has a 5 % market share. Diaper sizes 3 and 4 have the highest market shares (28~29 %) while size 1 has the lowest market share at 12 %.

Table 2 shows that Pampers is also the highest priced brand with Huggies being a close second and the store brand being the cheapest brand (at a price level that is on average 30 % lower than the two national brands).

Table 1 Sales, revenue, and market share summary statistics

Brands	Units Sold	Revenue	Purchase Shares
Store	937	9,559.57	0.052
Huggies	4812	87,201.72	0.470
Pampers	4368	88,716.71	0.478
Sizes	Units Sold	Revenue	Purchase Shares
1	1511	22,676.69	0.122
2	1500	28,143.83	0.151
3	2787	51,564.35	0.278
4	2716	53,709.00	0.289
5	1603	29,693.83	0.160

Table 3 is the switching matrix between consecutive purchases among different sizes of different brands.

Table 3 shows that repeat purchase of the same size of the same brands accounts for a large percentage of all purchases, while in the meantime, there are also a significant number of switches between different brands of the same size and switches between the adjacent sizes of different brands. In addition, there are also a significant number of switches across different brands in larger sizes of diapers. These switching patterns and variation across households in regard to purchase histories and switching patterns, as well as price variation over time, aid in identifying the cross-size learning effects, the duration-dependence effects in utility, and price sensitivities for smaller versus larger brands. A detailed discussion of how the individual learning parameters are identified in learning models is available in Crawford and Shum (2005), Erdem et al. (2008) and Ching et al. (2013).

Table 2 Marketing mix summary statistics

Brand	Size	Mean Price (dollars/piece)	Mean Feature (0/1)
Store	1	0.151	0.004
Huggies	1	0.223	0.008
Pampers	1	0.232	0.006
Store	2	0.178	0.003
Huggies	2	0.243	0.01
Pampers	2	0.244	0.007
Store	3	0.199	0.005
Huggies	3	0.281	0.009
Pampers	3	0.287	0.008
Store	4	0.220	0.006
Huggies	4	0.347	0.009
Pampers	4	0.314	0.008
Store	5	0.272	0.008
Huggies	5	0.378	0.014
Pampers	5	0.417	0.012

Table 3 Brand-size switching matrix

Brand	1 ST	1 HU	1 PA	2 ST	2 HU	2 PA	3 ST	3 HU	3 PA	4 ST	4 HU	4 PA	5 ST	5 HU	5 PA
ST	21	5	3	15	2	0	0	0	0	0	0	0	0	0	0
HU	2	63	4	3	33	2	0	0	0	0	0	0	0	0	0
PA	3	3	82	2	3	43	0	0	0	0	0	0	0	0	0
ST	1	0	1	15	1	0	5	1	1	0	0	0	0	0	0
HU	1	9	0	2	92	4	2	36	3	0	0	0	0	0	0
PA	1	0	7	0	2	47	5	3	20	0	0	0	0	0	0
ST	0	0	0	4	1	0	25	2	0	3	1	0	0	0	0
HU	0	0	0	3	25	2	4	132	15	6	54	3	0	0	0
PA	0	0	0	0	0	10	2	17	93	8	4	29	0	0	0
ST	0	0	0	0	0	0	2	5	0	4	4	2	4	1	0
HU	0	0	0	0	0	0	4	37	1	6	96	4	9	30	2
PA	0	0	0	0	0	0	5	2	21	7	7	71	7	4	22
ST	0	0	0	0	0	0	0	0	0	4	0	0	10	5	2
HU	0	0	0	0	0	0	0	0	0	3	19	4	5	83	6
PA	0	0	0	0	0	0	0	0	0	2	2	15	7	7	52

ST Store, *HU* Huggies, *PA* Pampers

Given that in this paper our focus is on learning spillover effects, identification of these deserves special attention. The spillover effect coefficients across sizes (κ) (Eq. 6) are functions of the standard deviation of experience variability (σ_ξ), the standard deviation of initial perception variability (σ_υ), and finally the initial correlations across perception errors (ρ's). The experience variability (σ_ξ^2) and the standard deviation of initial perception variability (σ_υ) are identified by the evolution of household choices over time and heterogeneity in that evolution (different households buy different brands over time). That is, these parameters are identified by the choices of households who buy and experience a brand versus the ones who do not, and by the subsequent choices of those who bought that brand. This identification is standard in the learning models estimated on scanner panel data in the literature (e.g., Erdem and Keane 1996). The initial correlation parameters across perception errors (ρ's) are also identified by the same data patterns, as well as by differences between unconditional brand switching probabilities versus brand switching probabilities conditional on switching to the adjacent size for each size.[20] These conditional brand switching percentages (conditional on a size switch) were 12.5, 25.6, 22.5, 20.8, 19.3 %, for sizes 1 through 5, respectively. These are higher than the corresponding unconditional brand switches: 11.1, 10.4, 15.4, 19.6, and 19.0 % for Size 1 through Size 5, respectively. We should stress here the following. First, the conditional brand switching percentages are higher than unconditional ones, consistent with the notion that people may be motivated doing strategic sampling when they are switching sizes. Second, the differences between conditional and unconditional probabilities, as well

[20] That is, count all brand switches when a (an adjacent) size switch was made and divide it by the total number of size switches between adjacent sizes irrespective of whether there was brand switching or not.

as conditional probabilities themselves, go down as sizes grow (except for size 1[21]), consistent with the notion that there is always more return on trial if one tries early on.[22] These data patterns help us identify our learning parameters. Additionally, to assess the fit of our model and also to further shed light on the identification of the initial correlations across perception errors (ρ's), we ran simulations where we set the ρ's to be zero or one. We discuss these simulations in section 4.3.

We also should stress here again that the standard learning models (e.g., Erdem and Keane 1996) imply that consumers will settle on small subset of brands once consumers are less uncertain about (a subset) of brands, *ceteris paribus.* Thus, standard leaning models would suggest that we should observe less switching over time as consumers get more experienced. In the diapers case, this would mean, for example, we would observe less switching for larger sizes than smaller sizes. However, if needs change (e.g., need to switch to a bigger size) and this may lead to a temporary increase in switching around the size change time, the smoothly declining switching pattern would not be observed, even if one would still expect overall there would be less switching as time passes by. Additionally, if consumers become more price elastic and/or sensitive when they learn more about the brands in the category, switching due to price promotions may increase, which may more than offset the declining switching due to learning. Indeed, when we calculated switching matrices for each size, we did *not* observe such a declining trend in switching. Instead, when we calculated the percentage of brand switches between brands in Table 3, we find the percentages of brand switching observations to be 11.1, 10.4, 15.4, 19.6, and 19.0 % for Size 1 through Size 5, respectively, indicating an *increasing* percentage of brand switches when size increases.[23] These switching observations help us also to identify the size-specific price parameters.

Finally, recall that the term $\mu_h(D_{h,t-1,k} - s_k D^2_{h,t-1,k})$ in the Eq. (1) captures duration dependence in a size. μ_h is identified by the duration consumers stay with a particular size whereas s_k is identified by the variation in duration for different sizes.

We would like to stress the fact that like in any other structural dynamic demand model functional form assumptions aid in identification in our estimation as well. Of course, state dependence cannot be distinguished non-parametrically from variation in

[21] When consumers are just starting with the newborn size 1, probably they are just learning about diapers and babies' diaper needs in general and not too motivated to start sampling a lot right away.

[22] We should also note here that most of the size switches occur to the adjacent bigger, rather than the adjacent lower size. Indeed, brand switching probabilities conditional on switching to the next bigger size are 12.5, 23.3, 21.6 % and 19.7 for sizes 1 through 4, respectively. When one counts only switches to the next bigger size, there is of course no such switch for size 5 as this is the biggest size.

[23] The increased number of switches in larger sizes would occur if the impact of increased price sensitivity & elasticity dominates the effect of diminished overall strategic sampling on brand switching. To again check data patterns, we categorized the switching observations into two groups: when a household switched to a different brand when the price of the brand switched to is at least 5 % lower than its mean price, we categorized the brand switch observations as brand switching due to price promotion. Otherwise, we classified the brand switch as a not price promotion related brand switching (which could be due to strategic trial or other reasons). The size-specific brand switching observations categorized as "price promotion related" yielded percentages of brand switching to be 44, 48, 51, 59 and 63 % for Size 1 through Size 5, respectively. That is, while 44 % of all size 1 brand switching was "price promotion related" (and 56 % was "non-promotion-related), 63 % of total size 5 brand switching was "price promotion-related". Thus, the data patterns suggest non-promotion related brand switches decline over time relative to price promotion related switches. Thus, consumers switch early on more for non-price related reasons (e.g., for strategic trial) while they will switch due to price variation in later periods.

behavior due to a completely general form of heterogeneity (Chamberlain 1984). Nor can learning behavior be distinguished non-parametrically from other mechanisms that may induce state dependence. Functional forms of both state dependence and heterogeneity must be constrained for the learning model to be identified. This is so for any non-linear dynamic model. From the perspective of a structural econometrician this is not a limitation – a model that simply specifies a very general form of state dependence is merely a statistical model that has no structural/behavioral interpretation. We do not wish to work with models that have extremely general forms of heterogeneity and/or state dependence. What we seek are parsimonious models that fit well and give insights into the data.

Before concluding this section we list briefly our formal identification restrictions here again: we set Q_{35}=1 (i.e., Store Brand Size 5 quality=1) and measure quality of other brands relative to this product. Furthermore, in the latent-class model, we set the quality weight parameter (w_k) in one of the segments to be 1 as commonly done in this literature (e.g., in a similar vein, Erdem (1998) fixes the variance of the heterogeneity distribution of the utility weights). Although there is no formal reason for this restriction, it is difficult to estimate the model without this restriction as the likelihood tends to be too flat. Keane (1992) calls this fragile identification since although a parameter may be formally identified, it may be impossible to estimate in practice because the likelihood is too flat; thus, the analysts may need a very (prohibitively) large number of observations in practice.

4 Empirical results

4.1 Model fit and model selection

Our model allows for heterogeneity in the price coefficient (α), duration weight (μ), utility weight on quality (w), risk coefficient (r), and the coefficient for inventory in the no-purchase utility (γ) so we must first choose the number of segments M. We estimated models with 1, 2, and 3 segments and report measures of fit in Table 4.

In the best-fitting model (forward-looking consumer choice with across-size learning, duration-dependence in utility, as well as size-specific quality and price parameters), increasing K from one to two improves AIC and BIC by 136 and 86 points, respectively; when we increase the number of segments from two to three, AIC and BIC increase by 80 and 23 points, respectively. While when we increase the number of segments further to four, the BIC did not improve. Based on these results, we decided to use the three-segment model for further analysis.

Comparing the fits of different models we estimated, we can also find out the relative importance of different components of our model. Compared to the model with only learning of brand-size quality (Model 1), modeling consumer learning across adjacent sizes (Model 2) improves the BIC by 164 points while adding duration dependence in the utility specification (Model 3) improves the BIC by 110 points. This shows the importance of accounting for both size learning and duration-dependence in utility in modeling consumer brand choice. The longer a consumer stays with a specific size of a brand, the more likely its utility will decrease. Table 4 also

Table 4 Model Selection

	Estimation Sample				Holdout Sample			
		One-Segment	Two-Segment	Three-Segment		One-Segment	Two-Segment	Three-Segment
Model 1: Learning Model with Myopic Consumers I	LL	−5432.3	−5381.25	−5308.37	LL	−4703.31	−4576.03	−4410.48
	AIC	10906.6	10814.5	10682.74	AIC	9448.62	9204.06	8886.96
	BIC	11056.04	10999.52	10917.58	BIC	9595.544	9385.967	9117.841
Model 2: Learning Model with Myopic Consumers II (Model 1+Learning across Adjacent Sizes)	LL	−5391.43	−5313.94	−5208.3	LL	−4675.97	−4520.14	−4307.37
	AIC	10832.86	10687.88	10490.6	AIC	9401.94	9100.28	8688.74
	BIC	11010.77	10901.37	10753.9	BIC	9576.85	9310.172	8947.607
Model 3: Learning Model with Myopic Consumers III (Model 2+Duration-dependent Parameters in the Utility)	LL	−5300.81	−5309.15	−5120.95	LL	−4582.31	−4358.37	−4118.49
	AIC	10663.62	10692.3	10329.9	AIC	9226.62	8790.74	8324.98
	BIC	10884.22	10955.6	10643.01	BIC	9443.509	9049.607	8632.822
Model 4: Learning Model with Forward-looking Consumers I (Model 3+Forward-looking Consumers)	LL	−5282.08	−5154.71	−5024.93	LL	−4555.22	−4314.92	−4051.93
	AIC	10626.16	10383.42	10137.86	AIC	9172.44	8703.84	8191.86
	BIC	10846.76	10646.72	10450.97	BIC	9389.329	8962.707	8499.702
Model 5: Learning Model with Myopic Consumers IV (Model 3+Size-specific Price Parameters)	LL	−5278.11	−5158.29	−5037.74	LL	−4540.84	−4301.33	−4028.85
	AIC	10622.22	10398.58	10175.48	AIC	9147.68	8684.66	8157.7
	BIC	10857.06	10690.35	10531.29	BIC	9378.561	8971.513	8507.52
Model 6: Learning Model with Forward-looking Consumers II (Model 5+Forward-looking Consumers) as well as (Model 4+Size-Specific Price Parameters)	LL	−5108.33	−5033.52	**−4985.75**	LL	−4408.87	−4193.4	**−4018.77**
	AIC	10282.66	10147.04	**10067.5**	AIC	8883.74	8466.8	**8133.54**
	BIC	10517.5	10431.69	**10409.08**	BIC	9114.621	8746.656	**8469.367**

* Calibration sample: Number of observations=9102. Number of households=119. Holdout sample: Number of observations=8074. Number of households=106. Number of weeks=78

** Note: AIC=−2*Log-likelihood+2*# of parameters; BIC=−2*Log-likelihood+# of parameters*ln (# of observations)

shows that the 3-segment myopic learning model (Model 3) has a log-likelihood that is 293 points worse than the 3-segment model with forward-looking consumers (Model 4), and the 3-segment myopic learning model with size-specific parameters (Model 5) has a log-likelihood that is 52 points worse than the 3-segment model with size-specific parameters and forward-looking consumers (Model 6). Thus, the forward-looking aspect of the model (i.e., strategic trial purchases) is important.

Comparing Model 4 (forward-looking learning model without size-specific price parameters) with Model 6 (forward-looking learning model with size-specific price parameters), we find Model 6 improves the BIC by 41 points. This implies that for the consumers in our diaper purchase data, the price sensitivities change with time. Finally, the best fitting model (Model 6) has adjacent size learning spill-over effects, duration-dependent size utility and forward-looking consumers (as well as price sensitivities that differ for small versus large sizes). As we previously indicated, this model allows for the possibility that consumers may temporarily increase strategic trial around the time they switch sizes.

Table 5 Estimates of the homogeneous part of model 5 with 3 Segments

Parameter		Estimates	Std error
Size 1 Quality	Store	**−0.891**	0.16
	Huggies	**0.051**	0.02
	Pampers	**0.045**	0.02
Size 2 Quality	Store	**0.011**	0.01
	Huggies	**0.068**	0.02
	Pampers	**0.084**	0.03
Size 3 Quality	Store	**0.107**	0.05
	Huggies	**0.135**	0.03
	Pampers	**0.165**	0.03
Size 4 Quality	Store	**0.025**	0.01
	Huggies	**0.041**	0.02
	Pampers	**0.025**	0.01
Size 5 Quality	Huggies	**0.077**	0.03
	Pampers	**0.098**	0.04
Duration_size1 (s_1)		0.874	0.83
Duration_size2 (s_2)		**0.742**	0.05
Duration_size3 (s_3)		**0.633**	0.03
Duration_size4 (s_4)		**0.558**	0.04
Duration_size5 (s_5)		**0.302**	0.05
Size Correlation Between 1 and 2 (ρ_{12})		**0.457**	0.05
Size Correlation Between 2 and 3 (ρ_{23})		**0.445**	0.04
Size Correlation Between 3 and 4 (ρ_{34})		**0.459**	0.04
Size Correlation Between 4 and 5 (ρ_{45})		**0.512**	0.05
Standard Deviation of Initial Perception Variability (σ_υ):			
Intercept (λ_0)		**9.865**	0.85
Number of purchases in initialization period (λ_1)		**−1.862**	0.78
Standard Deviation of Experience Variability (σ_ξ)		**1.435**	0.09
Standard Deviation of Experience Variability for across size learning (σ_η)		**14.016**	1.33

Note: Parameters highlighted in bold are 95 % statistically significant; parameters highlighted in bold and italic are 90 % statistically significant

For validation purpose, we tested the proposed model on the holdout sample of 106 households. Parameters corresponding to Models 1 through 6 were used to compute the likelihood functions values for the holdout households. We find the proposed learning model (with size spill-over effect, duration-dependent utility parameters, and size-specific price parameters) with forward-looking consumers continues to be the best-fitting model for the hold-sample based on the AIC and BIC criteria. This supports the predictive validity of the proposed model.

4.2 Parameter estimates

We report parameter estimates for our preferred (three-segment) model (Model 6) in Tables 5 and 6.

Table 6 Estimates of the Heterogeneous Part of Model 5 with 3 Segments

Parameter	Segment 1		Segment 2		Segment 3	
	Estimates	Std error	Estimates	Std error	Estimates	Std error
Price ($\beta_{h\_small\_size}$)	**−0.410**	0.20	**−3.951**	1.53	2.328	1.80
Price ($\beta_{h\_large\_size}$)	**−1.428**	0.43	**−8.587**	1.46	0.117	1.98
Duration (μ_h)	**0.010**	0.00	**0.004**	0.00	**0.269**	0.07
Inventory (γ_h)	**0.034**	0.01	**0.028**	0.01	**0.002**	0.00
Utility Weight (w_h)	**6.931**	1.28	**7.017**	0.98	1	
Risk Coefficient (r_h)	**−0.847**	0.42	**−6.641**	1.16	**−11.257**	1.16
Segment Size Weight	**−1.532**	0.07	**−0.277**	0.02		
Size of Segment 1	22 %					
Size of Segment 2	76 %					
Size of Segment 3	3 %					

Note: Parameters highlighted in bold are 95 % statistically significant; parameters highlighted in bold and italic are 90 % statistically significant

Table 5 lists the estimates that are homogenous across segments. The quality estimates are all statistically significant.[24] We find the estimates of the four initial correlation coefficients (ρ) across different sizes to be all positive (between 0.445 and 0.512) and significant. These are the correlation coefficients for the initial perception errors across sizes; therefore, positive estimates indicate that consumers learn through consumption across sizes. Thus, use experience with a specific size gives information about the quality of its two adjacent sizes as well.

The two parameters of the prior quality perception standard deviation (σ_υ) specification are 9.865 for the intercept (λ_0) and −1.862 for the number of purchases in the initialization period coefficient (λ_1). This suggests that quality uncertainty exists and it is a function of number of past purchases in the category. The estimate of the standard deviation of the experience variability (σ_ξ) (capturing the noise in the consumption experience as a source of quality information) is statistically significant and 1.435. Thus, use experience provides (noisy) information about quality. The corresponding variability (σ_η) for across size experience signals is statistically significant as well and is 14.016, suggesting that use experience associated with a brand-size provides noisy information about the adjacent sizes of the same brand. As one would expect, the estimate of (σ_ξ) is lower than that of (σ_η) since quality information about a specific brand size obtained through use experience is expected to be less noisy than quality information obtained about that specific brand size through experience of an adjacent size of the same brand.

Table 6 lists the estimates that differ by segments. Size duration weight parameter μ_h is positive for all three types of consumers. The estimates of the five location parameters s_k are also positive and bound between 0 and 1 (between 0.302 and 0.742), and

[24] The "true qualities" can be different across segments if there is a baby-diaper match issue. The same issue holds in many other categories as well though and it is not feasible to estimate a dynamic structural model that allows true qualities varying by households. Furthermore, even if the match issue exists, there is no reason to believe that not modeling it would bias the results in a systematic way.

they are also highly significant except for s_1 (location parameters s_k are homogenous across segments and are given in Table 5). Given the quadratic specification of duration-dependence in our utility function, the above estimates imply that the utility increases with the size duration for as long as $D_{hkt} < 1/s_k$, and thereafter it starts to decrease. Interestingly, the estimates s_k are larger for smaller sizes and become smaller for larger sizes. This implies that consumer utilities decrease with the time spent on a size, more quickly so for smaller diaper sizes than for larger diaper sizes.

The price coefficients are found to be negative and significant for two of the three segments, which account for 97 % of the households in our sample, in the full model (Model 6). We find the consumer's price coefficients are higher for the larger size than for the smaller size. In segment 1 (22 % of the sample), the price coefficient is −1.428 for the larger size and −0.410 for the smaller size; while in segment 2, the largest segment (76 % of the sample), the price coefficient is −8.587 for the larger size and −3.951 for the smaller size. The price coefficient estimates from both segments imply that first-time parents are less price-sensitive when they start buying diapers (small size) for their children, while they become more price-sensitive after they buy the diapers for a while and start buying larger sizes of diapers. Finally, we also find the consumer's quality weight coefficients are positive and significant for all three segments.

It might be useful to compare our results with those results in Heilman et al. (2000). Different from Heilman et al. (2000) who found price coefficient (sensitivity) to be declining in use experience (cumulative previous purchases), we find more experienced consumers (buyers of larger sizes) to be more sensitive that buyers of smaller sizes. In their study, the authors do not have a priori prediction that the price sensitivity will become higher or lower. We think that the finding of increasing price sensitivity when households' babies grow older and parents accumulate more information is intuitively more appealing than the reverse result. As previously discussed, for example, first time parents may not have a very good sense of their total diaper expenses when they first enter the market and figure out their expenses better when time passes, which may lead for the marginal utility of income change. Marginal utility of income may also increase due to increased expenses of growing kids.

To explore the effects of learning and reduced uncertainty on price elasticities, we run price experiments in Section 5 and find that price elasticities are also larger for larger brands. Both set of results (increased price elasticities and price sensitivities) also explain better the fact that store brands market shares for larger size diapers are higher than those of smaller size diapers, which is the case both in our and Heilman et al. (2000)'s data.

Overall, different from the descriptive results from Heilman et al. (2000) and many papers in the structural dynamic consumer choice literature, the evidence we obtain from the parameter estimates suggests that in the diapers category, consumers do not settle on one brand or a very small set of brands in the diapers category due to learning effects (also labeled as familiarity effect in the descriptive model literature). This is because price sensitivities and elasticities increase when consumer uncertainty decreases over time. Furthermore, our model implies that strategic sampling may go up temporarily when the household is ready to switch sizes.[25]

[25] This implication is consistent with data that show that there is indeed more brand switching conditional on size switching.

Table 7 Effects of initial perception error correlations on choice probabilities

		Store	Huggies	Pampers
ρ_{12}=0	Store	−0.03	0.01	0.05
	Huggies	0.01	−0.08	0.05
	Pampers	0.01	0.04	−0.09
ρ_{23}=0	Store	−0.05	0.02	0.02
	Huggies	0.01	−0.05	0.06
	Pampers	0.01	0.04	−0.08
ρ_{34}=0	Store	−0.03	0.01	0.02
	Huggies	0.01	−0.07	0.04
	Pampers	0.02	0.05	−0.08
ρ_{45}=0	Store	−0.01	0.01	0.01
	Huggies	0.01	−0.05	0.02
	Pampers	0.01	0.03	−0.04

Note: Entries in the cell are the choice probabilities calculated from the proposed model by allowing one of the correlation coefficients (ρ's) to be zero

4.3 Model fit simulations

We run three simulations[26] involving the initial correlations of perception errors across adjacent sizes (ρ's). In Table 7, we sequentially set the values of ρ's of the corresponding two sizes to be zero and investigate its effect on the choice probabilities of the brand under consideration and competing brands.

We find when the value of ρ is set to zero, the choice probability for the brand under consideration goes down while the competing brands' choice probabilities go up. This is consistent with our finding of positive spillover effects of brand-size learning across adjacent sizes of the same brand. We also find the decreases in choice probabilities are much larger for Pampers and Huggies than for the store brand, indicating a much higher positive spillover effect for the two national brands.

In Table 8, we first calculate the base choice probabilities of different brand-sizes under the scenario that all the correlation coefficients, ρ's, are zero. Then we sequentially turn on the ρ's using our estimates and recalculate the choice probabilities. We find, with few exceptions, that an increase of size correlation coefficient from zero to positive values leads to higher choice probabilities for both adjacent sizes.

Finally, we ran simulations with initial correlations between size perception errors (ρ's) to be either zero or one to assess the impact of this manipulation on brand switching probabilities conditional on size switch. Table 9 reports the results.

Table 9 reports unconditional brand switching probabilities and brand switching probabilities conditional on size switch observed in the data and predicted by our estimated model, as well as conditional probabilities predicted if one sets initial correlations to zero or one. As it can be seen Table 9, setting these initial

[26] In the model fit simulations and counterfactual analyses, we first draw the parameter estimates from their joint distribution, then we solve for choice shares for each draw, and lastly calculate the mean choice shares across the draws.

Table 8 Effects of initial perception error correlations on choice probabilities

Holding all other ρ's to be 0, when:	Store		Huggies		Pampers	
	size k	size $k+1$	size k	size $k+1$	size k	size $k+1$
$\rho_{12}\neq0$	0.0002	0.0014	0.0040	0.0005	0.0052	0.0014
$\rho_{23}\neq0$	0.0033	0.0006	0.0053	0.0011	0.0034	0.0002
$\rho_{34}\neq0$	0.0003	−0.0001	0.0053	0.0053	0.0059	−0.0001
$\rho_{45}\neq0$	0.0003	0.0003	0.0046	0.0018	0.0005	0.0041

k=1,2,3,4. Entries in the cell are the choice probabilities calculated from the full models by allowing one of the correlation coefficients ($\rho_{k,k+1}$) to be non-zero, while holding other ρ's to be zeros.

correlations (ρ's) to zero overestimates the conditional probabilities whereas setting these to one underestimates the conditional probabilities. Thus, when initial perceptions are correlated across sizes (potential for learning across sizes) but when one ignores such correlation, one would get higher switching probabilities when consumers switch sizes since information a household has about one size has no information value for the other size and hence there is more a need for learning and less state dependence when one switches a size. In the other extreme, when the correlation is one, consumers thinks that the brand offers the same quality across sizes so there is no information loss across sizes and hence the conditional switching probabilities are underestimated.

5 Policy experiments

We run two policy experiments (using the estimates from the preferred model, i.e., Model 6) to explore the implications of our proposed and estimated model. We should note that in these simulations we assume no competitor response so this is a partial equilibrium analysis.

In Table 10, we simulate a 10 % price cut for the three brands that last for 1 week, and then calculate the cumulative change in consumer's choice probability over that week and the next 9 weeks. To distinguish the difference in consumer responses to

Table 9 Unconditional brand switching probabilities and brand switching probabilities conditional on size switching in the data, predicted by our model and predicted when ρ's are set to zero or one

	Observed		Predicted Based on Estimates		Predicted by setting ρ =0	Predicted by setting ρ =1
	unconditional	conditional	unconditional	conditional	conditional	conditional
Size 1	11.1 %	12.5 %	10.4 %	12.8 %	15.2 %	11.8 %
Size 2	10.4 %	25.6 %	10.8 %	24.3 %	27.9 %	21.7 %
Size 3	15.4 %	22.5 %	14.4 %	23.4 %	25.1 %	21.3 %
Size 4	19.6 %	20.8 %	20.9 %	21.5 %	24.3 %	20.4 %
Size 5	19.0 %	19.3 %	20.1 %	20.7 %	23.7 %	19.0 %

Table 10 Effects of Price Cuts on Choice Probability for Different Brands and Sizes

Cut prices of the small size of the three brands by 10 %				
		Price Elasticity		
		Store	Huggies	Pampers
Temporary 10 % price cut in	Store	0.337*	−0.050	−0.041
	Huggies	−0.141	0.308	−0.150
	Pampers	−0.146	−0.142	0.289
Cut prices of the large size of the three brands by 10 %				
		Price Elasticity		
		Store	Huggies	Pampers
Temporary 10 % price cut in	Store	0.428	−0.124	−0.127
	Huggies	−0.176	0.355	−0.152
	Pampers	−0.186	−0.149	0.330

*The probability of a consumer choosing store brand will increase by 0.337 given a 10 % decrease in price

price promotion of small and large sizes, we do two simulation exercises, one each size, and compare the results.

We have two important findings: 1) Consumer's own price elasticity for the large size is higher than that for the smaller size. For example, after a 10 % price cut, the own price elasticity for the large size of the store brand (the economy brand) is roughly 4.3, while it is 3.4 for the small size; similar results are found for the premium brands: Pampers and Huggies; 2) Consumer's cross price elasticity for the large size is higher than that for the smaller size. Here, the most interesting finding is for the store brand. We find, after a 10 % cut of the store brand's price, the cross elasticities for Pampers and Huggies with respect to the store brand are roughly −1.24 and −1.27 for the large size, while they are roughly −0.41 and −0.50 for the small size. Interestingly, the cross price elasticities for the store brand (economy brand) with respect to Pampers and Huggies are also larger for the large size, but the magnitude of increase is much smaller than those for Pampers and Huggies with respect to the store brand. In other words, when the store brand is offering a price cut of the large size of diapers, consumers are more likely to switch from premium (high quality) brands to economy (low quality) brands than they were when buying smaller brands.[27]

Second, we conduct policy experiments that examine the impact free samples have on consumer choices. The free sample implies that a consumer gets the chance to try a brand for free, and the simulations conducted assume that all the consumers get the free sample and everybody uses (consumes) it. These free sample policy experiments can further shed light on the cross-size learning effects on consumer choice. In addition, we can also investigate whether providing free samples will

[27] Additionally, we also repeated the price simulations with the parameters obtained from a model without size-specific price coefficients (this model differed from ours only by the fact that price coefficients were not size specific). We obtained similar results and price elasticities are still larger for larger sizes but compared to the model with size-varying price parameters, this effect is dampened a bit as one would expect. Thus, price elasticties are larger for larger sizes even when does not allow size specific price coefficients but the effect is more pronounced when there are size specific price coefficients.

lead to more strategic sampling from consumers on a particular brand as compared to other brands.

Tables 11 report the actual frequency of purchases for the 15 brand-sizes and baseline simulation results for 25 purchase occasions.

The *observed purchase frequency* row reports the numbers of households that bought a given brand-size for 25 purchase occasions. The baseline simulation row reports the predicted numbers of purchases for different brand-sizes obtained from the proposed model (Model 6) using the estimated parameters. As can be seen from these two tables, baseline simulation results approximate the observed purchase frequencies well, especially for frequently bought brand-sizes.

To assess the impact of alternative strategy changes, the baseline simulation results (Table 11) must be compared with post-intervention (the distribution of the free sample) figures. In Table 12, we report the post-intervention cumulative changes (in %) in purchase frequency after providing one unit of free sample of the small or large brand-size of diapers to each household at the end of the first week.

The results in Table 12 indicate that for our proposed model, free samples of the small size (sizes 1~3) provided by a brand before the second purchase occasion increase sales of that brand for sizes 1~3 and their adjacent size: size 4. Similarly, free samples of the large size (sizes 4~5) provided by a brand also increase sales of that brand for sizes 4~5 and their adjacent size: size 3. This could be due to the existence of both spillover effect (across sizes) and duration-dependent utilities. Overall, national brands benefit more from free sampling than the store brand. Also providing free samples of smaller sizes attracts more brand switchers than providing free samples of larger sizes. Interestingly, we also find for store brand, the gain from providing free sample of smaller size versus larger size is bigger.

6 Discussion and conclusions

We estimated for the first time in literature a forward-looking structural dynamic learning model where the need for learning and strategic sampling may increase periodically, there are spill-over learning effects across the different versions of a brand but not all prior information about the brand is retained when consumers migrate across versions of products. We also showed that price elasticities may increase as consumer uncertainty drops. Our estimation results show: 1) consumer experience of a particular size of a particular brand serves as a quality signal of the other (adjacent) sizes of the same brand, 2) consumer brand-size preference is duration dependent and it first increases and then

Table 11 Observed and simulated purchase frequency

Brand	Store					Huggies					Pampers				
Size	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Observed	13	20	46	12	23	31	62	129	87	55	50	47	77	79	64
Baseline Simulation	11	20	49	9	19	32	60	129	81	55	53	54	76	80	67

Table 12 Purchase Frequency from Free Sample Experiments

	Brand	Change in Own and Adjacent Sizes Between Free Sample and Baseline Simulation		
		Store	Huggies	Pampers
Free Sample in Size 1-3	Store	30.4 %	−7.0 %	−5.7 %
	Huggies	−20.2 %	19.3 %	−7.2 %
	Pampers	−22.5 %	−8.3 %	20.7 %
Free Sample in Size 4-5	Store	20.8 %	−3.0 %	−4.0 %
	Huggies	−18.8 %	15.8 %	−8.5 %
	Pampers	−22.1 %	−7.9 %	16.6 %

decreases with the time that consumers stay with a particular brand-size, and 3) consumer price elasticities are higher for larger size diapers (as consumers learn more about brand qualities when their babies grow older, price elasticities increase). We also found the price sensitivities to be different for smaller versus larger sizes.

Our policy experiment results include the finding that when faced with a price promotion, consumers are more likely to switch from premium (national) brands to economy (store) brands when buying larger sizes than smaller sizes. Our free sampling simulation analysis indicates that while free sampling is overall more beneficial for national brands and providing free samples of smaller sizes is better than providing free samples of larger sizes for all brands, the differential gain between smaller size versus larger size free samples is bigger for store brands.

These combined results suggest a number of managerial implications. First, our results show that consumers who just enter the diaper market are less price sensitive, while they become more price-sensitive when their babies grow (and they gain more experience). The results from our study could help managers do a better job at developing promotion strategies to different consumer segments. More specifically, national brands could focus on providing free samples to consumers who are new to the market, and focus on price promotions to consumers with more category experience. We show also that consumer preference for a brand is duration-dependent and it could decline when their needs change (e.g., the baby is growing out of a size). Managers could develop promotion strategies by tracking a consumer's purchase history and giving free samples (or coupons) around the time that her needs change. Second, given increased price sensitivities over time, store (or private) label managers could determine the optimal time of pursuing a consumer aggressively. National brands, on the other hand, may try to reverse this trend by adding new features to larger size diapers.

Going beyond the diapers category, the general implications of this research include the fact that firms need to be aware of the timing of consumers increased motivation for sampling (e.g., in brand relaunches, the timing will be the same for all households whereas in other cases, specific demographics changes will lead household-level specific timing implications), as well as systematically evolving consumer sensitivities. Varying-parameter models (e.g., Mela et al. 1997) have captured stochastically evolving preferences or preferences that evolve as a function of marketing mix and this research shows a systematic evolution of such sensitivities over time in markets where there is quality uncertainty.

Acknowledgment Hai Che acknowledges the financial support from the National Natural Science Foundation of China (71428004).

References

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

Anderson, S. T., Kellogg, R., Langer A., Sallee J. M. (2013). The Intergenerational Transmission of Automobile Brand Preferences: Empirical Evidence and Implications for Firm Strategy. NBER Working Paper 19535. http://www.nber.org/papers/w19535.

Bucklin, R. E., & Gupta, S. (1992). Brand choice, purchase incidence, and segmentation: an integrated modeling approach. *Journal of Marketing Research, 29*, 201–215.

Chamberlain, G. (1984). Panel Data. In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics*, Chapter 22, Vol. 2, (pp. 1247–1318). North Holland Press.

Chan, T., Narasimhan, C., Xie Y. (2013). Treatment effectiveness and side-effects: a model of physician learning. *Management Science*

Che, H., Sudhir, K., & Seetharaman, P. B. (2007). Bounded rationality in pricing under state dependent demand: do firms look ahead? How far ahead? *Journal of Marketing Research, 44*(3), 434–449.

Ching, A., Erdem, T., & Keane, M. (2013). Learning models: an assessment of progress, challenges and new developments. *Marketing Science, 32*(6), 913–938.

Ching, A., Erdem T., Keane, M. (2014). A simple approach to estimatethe roles of learning, inventory and experimentation in consumer choice. *Journal of Choice Modeling.*

Chung, D. J., Steenburgh, T., & Sudhir, K. (2014). Do bonuses enhance sales productivity? A dynamic structural analysis of bonus-based compensation plans. *Marketing Science, 33*(2), 165–187.

Coscelli, A., & Shum, M. (2004). An empirical model of learning and patient spillover in new drug entry. *Journal of Econometrics, 122*(2), 213–246.

Crawford, G. S., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica, 73*(4), 1137–1173.

DeGroot, M. H. (1970). Optimal statistical decisions. McGraw-Hill.

Dickstein, M. J. (2011). Efficient provision of experience goods: evidence from antidepressant choice. Working paper. Harvard Business School.

Dubé, J. P., Hitsch, G., & Rossi, P. (2009). Do switching costs make markets less competitive? *Journal of Marketing Research, 46*(4), 435–445.

Erdem, T. (1998). An empirical analysis of umbrella branding. *Journal of Marketing Research, 35*(3), 339–351.

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: capturing choice dynamics in turbulent consumer goods markets. *Marketing Science, 15*, 1–21.

Erdem, T., Keane, M. P., Sabri Öncü, T., & Strebel, J. (2005). Learning about computers: an analysis of information search and technology choice. *Quantitative Marketing and Economics, 3*(3), 207–246.

Erdem, T., Keane, M., & Sun, B. (2008). A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science, 27*(6), 1111–1129.

Goettler, R., & Gordon, B. R. (2011). Does AMD spur Intel to innovate more? *Journal of Political Economy, 119*(6), 1141–1200.

Gowrisankaran, G., & Rysman, M. (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy, 120*, 1173–1219.

Gupta, S., Kaul, A., & Wittink, D. R. (1996). Do household scanner panel data provide representative inferences from brand choices: a comparison with store data. *Journal of Marketing Research, 33*(November), 383–399.

Hartmann, W. R. (2006). Intertemporal effects of consumption and their implications for demand elasticity estimates. *Quantitative Marketing and Economics, 4*(4), 325–349.

Hartmann, W. R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science, 29*(4), 585–601.

Heckman, J. J. (1981). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in labor markets* (pp. 91–139). Chicago: Chicago University Press.
Heilman, C. M., Bowman, D., & Wright, G. P. (2000). The evolution of brand preferences and choice behaviors of consumers new to a market. *Journal of Marketing Research, 37*(May), 139–155.
Huang, Y. (2015). Learning by Doing and the Demand for Advanced Products. Working Paper.
Keane, M. P. (1992). A note on identification in the multinomial probit model. *Journal of Business & Economic Statistics, 10*(2), 193–200.
Keane, M. P., & Wolpin, K. I. (1994). Solution and estimation of dynamic programming models by simulation. *Review of Economics and Statistics, 76*(4), 648–672.
Li, S., Sun, B., & Wilcox, R. T. (2005). Cross-selling sequentially ordered products: an application to consumer banking services. *Journal of Marketing Research, 42*(May), 233–239.
McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of Econometrics* (pp. 105–142). New York: Academic.
Mela, C. F., Gupta, S., & Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research, 34*, 248–261.
Osborne, M. (2011). Consumer learning, switching costs, and heterogeneity: a structural examination. *Quantitative Marketing and Economics, 9*(1), 25–70.
Rust, J. (1987). Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica, 55*(5), 999–1033.
Shin, S., Misra, S., & Horsky, D. (2012). Disentangling preferences and learning in brand choice models. *Marketing Science, 1*, 115–137.
Sun, B. (2005). Promotion effect on endogenous consumption. *Marketing Science, 24*(3), 430–443.
Yao, S., Mela, C. F., Chiang, J., & Chen, Y. (2012). Determining Consumers' discount rates with field studies. *Journal of Marketing Research, 49*(6), 822–841.