



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Targeting Mr. or Mrs. Smith: Modeling and Leveraging Intrahousehold Heterogeneity in Brand Choice Behavior

Hernán A. Bruno, Javier Cebollada, Pradeep K. Chintagunta

To cite this article:

Hernán A. Bruno, Javier Cebollada, Pradeep K. Chintagunta (2018) Targeting Mr. or Mrs. Smith: Modeling and Leveraging Intrahousehold Heterogeneity in Brand Choice Behavior. Marketing Science 37(4):631-648. <https://doi.org/10.1287/mksc.2018.1088>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2018, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Targeting Mr. or Mrs. Smith: Modeling and Leveraging Intrahousehold Heterogeneity in Brand Choice Behavior

Hernán A. Bruno,^a Javier Cebollada,^b Pradeep K. Chintagunta^c

^a University of Cologne, 50923 Cologne, Germany; ^b Institute for Advanced Research in Business and Economics, Public University of Navarra, 31006 Pamplona, Spain; ^c Booth School of Business, University of Chicago, Chicago, Illinois 60637

Contact: hernan.bruno@wiso.uni-koeln.de,  <http://orcid.org/0000-0003-3677-5492> (HAB); cebollada@unavarra.es (JC); pradeep.chintagunta@chicagobooth.edu (PKC)

Received: January 10, 2017

Revised: September 5, 2017;
November 30, 2017

Accepted: December 15, 2017

Published Online in Articles in Advance:
July 12, 2018

<https://doi.org/10.1287/mksc.2018.1088>

Copyright: © 2018 INFORMS

Abstract. We develop a hierarchical choice model to account for the choice utility heterogeneity of individual shoppers that belong to the same household. Our model allows us to measure how much variability in purchase behavior exists among individuals in a household, and to compare this to the variability that exists across households. Because of the presence of multiple shoppers from the same household, we also extend the concept of household-level state dependence to consider state dependence at the individual level. We apply our model to five different grocery categories. We find that the intrahousehold heterogeneity in estimated brand intercepts and (to a lesser extent) price sensitivities is about 20%–30% of the interhousehold heterogeneity in these parameters. However, with promotion sensitivities, we find intrahousehold heterogeneity, in most cases, to be as large as interhousehold heterogeneity. Our state dependence results show that past brands purchased by an individual have a much stronger influence on subsequent purchases than those purchased by anyone in the household. We use our estimated utility parameters to compare the expected profitability of promotions targeted at the individual rather than at the household and find substantial (more than 50%) improvements in the incremental revenue of supermarket promotions.

History: Avi Goldfarb served as the senior editor and Harald van Heerde served as associate editor for this article.

Funding: This project was partially supported by the Spanish Ministry of Science [Research Project ECO2015-65393-R] and the Kilts Center of Marketing at the Booth School of Business, University of Chicago.

Supplemental Material: Data and the online appendix are available at <https://doi.org/10.1287/mksc.2018.1088>.

Keywords: choice model • targeting • heterogeneity • Bayesian estimation

1. Introduction

Studies of consumer purchase and choice behavior using scanner panel data take the household as the most disaggregated unit of analysis. Using discrete choice models of utility-maximizing behavior and with data on households' choices over time, these studies estimate household-level utility parameters that rationalize the observed choices made by households. These utility parameters can be used as the basis for targeting price promotions to individual households. However, modeling a household as a monolithic unit with one set of household-level utility parameters ignores that many households are composed of more than one individual. If these individuals differ in the utility parameters that rationalize their individual choices, inferences made from aggregated purchase behavior at the household level are unlikely to accurately reflect the behaviors of individuals. As a result, targeting policies based on household-level parameters as in Rossi et al. (1996) can lead to suboptimal marketing decisions.

Consider a household with two individuals (W and M) who make shopping trips with similar

regularity. Both W and M make “household” purchases in a category with two brands (A and B). A data set aggregated at the household level would show the sequence of purchases of the household. Let us further assume that the observed choices made by the individual members are such that W buys only brand A, while M buys only brand B. If the price of the brand W (M) buys, that is, A (B), is high at one particular occasion, she would rather not purchase than switch to the other brand, even if the price of that brand was low. Seen as a household, we would observe a sequence of purchases for both brands, and little influence of price on which brand gets chosen. A choice model estimated at the household level would reveal a lower price sensitivity than that of either W or M.

The aggregation at the household level affects the inferences of state dependence of choices as well (Gupta and Steckel 1993). Suppose W and M show positive state dependence of demand when they choose from a category, buying the same brand as before, and occasionally switching to another brand. If W and M alternate trips to the store, looking at these purchases

at the household level of aggregation may indicate weak or nonexistent state dependence. Since promotions are sometimes used to incentivize customers to switch brands and start a streak of purchases, observing the aggregate pattern may lead to incorrectly inferring that this household has weak state-dependent behavior, which might lead to the wrong pricing and promotional decisions.

Understanding the behavior of individuals within the context of a household is more relevant today than ever before because the digitization of information is enabling companies to not just track individual customer purchases but also link individual customers to other individuals that are related to them, either in their family or their social group. iTunes allows family members to share media purchases, and Amazon Prime services are available to family members. Many loyalty programs have records of which individuals belong to the same household (e.g., Tesco in the United Kingdom). More generally, intrahousehold interactions play a role in broader contexts such as transportation (e.g., Bhat and Pendyala 2005), resource allocation (Browning and Chiappori 1998), and spousal control (Ladner 2009).

In this paper, we study a household panel for which we observe the choices of individual household members (for a subset of households). Our main objective is to characterize heterogeneity in the utility parameters for brand, price, and promotion sensitivities *within* a household and its implications for firm targeting policies. Since we observe only the purchases made by individuals, we can only infer the individual utility parameters that rationalize the choices made by those individuals in a household. However, the choice utilities are the manifestation of not only that individual's preferences but also the preferences of other household members who may influence the choice decision. To make this distinction clear, we use the term "utility parameters" in the traditional sense used in choice modeling as parameters that explain the observed purchases by an individual or a household. We reserve the term "preference" to refer to the underlying individual consumption preferences. To the extent that each member of a household takes into account the preferences of the other members when making her choices, we cannot identify those preferences from individual choice data alone. However, since a firm can target based only on the behaviors it observes, the utility parameters we estimate using the choice data are appropriate for targeting individuals in a household.

Our modeling approach extends the traditional hierarchical Bayesian model of discrete choice (Rossi and Allenby 1993) to data where we observe which purchase belongs to which family member for a subset of households. As we cannot consider the choices made by one household member as independent of the choices of the others, we introduce a further hierarchy

in the estimation. We obtain a household-level vector of utility parameters, as has been done in other studies in the literature, and the individual members' parameter vectors as multivariate normally distributed deviations from the household parameter vector. We show how state dependence can be included in such a model; in particular, we account for state-dependent behavior at both the household and individual levels.

We perform our empirical analysis using data from a major European retailer in five frequently purchased categories (some results are only shown for ground coffee, our focal category). In our analyses and discussion, we use the term *interhousehold heterogeneity* to refer to the variability in choice parameters across households. The term *intrahousehold heterogeneity* refers to the variability across individuals *that belong to the same household*.¹

We provide a comparison of our model's fit and predictive ability with a model that uses individual data without considering the household membership of the individuals, and with a household-level model that ignores information on the individual members. The improvement in fit shows that the hierarchical model is flexible enough to accommodate individual variability while avoiding overfitting the model to individuals with few purchases.

Our results indicate that intrahousehold heterogeneity in estimated brand intercepts and promotion and price sensitivity cannot be ignored. Even though the intrahousehold heterogeneity is smaller than the interhousehold heterogeneity, it is within the same order of magnitude. For the brand intercepts, the standard deviations of the intrahousehold heterogeneity distribution are, on average, about 20%–30% of those of the interhousehold heterogeneity. In the case of price sensitivity, intrahousehold heterogeneity is closer in magnitude to—but still smaller than—interhousehold heterogeneity. In terms of sensitivity to promotions, we find that intrahousehold variability is of a similar magnitude and sometimes even larger than interhousehold heterogeneity. Another way of stating these results is that household members are likely to have more similar choice utility parameters related to brand preferences than to price or promotion sensitivities. Our results are consistent with the empirical finding in the behavioral marketing literature that people in a relationship do not necessarily share the same attitudes toward spending. Furthermore, they can often have opposite tendencies to spend—spendthrifts versus tightwads (Rick et al. 2011).

To understand the usefulness of individual versus household promotion targeting, we turn to a targeting exercise à la Rossi et al. (1996). Since there could be heterogeneity within the household, it would be appropriate to target the household member most responsive to the promotion. First, we show the potential benefits

of such a targeting policy. In the case that one household member is less likely to make a shopping trip, an efficient promotion would be one that also accounts for the individual within the household making the shopping trip. We augment our brand choice model and the targeting exercise with a model that computes the probability of a household member making the shopping trip. We use a hierarchical logit specification to infer the probability that one of the household members goes on the shopping trip as a function of their intrinsic shopping rates and other covariates. We then show the benefits of leveraging this additional information in our targeting. Our objective here is not to propose an “integrated” model of household member purchase behavior. Rather, the idea is to augment our brand choice model to improve targeting efficiency.

In sum, this paper makes the following contributions: we develop a choice model that uses purchase data disaggregated at the individual level and incorporates information about the individual membership to a household. The proposed model and estimation recover utility parameters from choices of individuals belonging to the same household, while allowing them to be correlated via a Bayesian hierarchy. The model also helps distinguish between individual and household state dependence. We apply our model to five different grocery categories. We find intrahousehold heterogeneity in estimated brand intercepts, price sensitivities, and responsiveness to promotions. Furthermore, intrahousehold heterogeneity in promotion sensitivities is the highest compared to the amount of interhousehold heterogeneity. The model shows a better fit and predictive ability than a model using individual data without considering the household membership of the individuals or a household-level model that ignores information on the individual members. We demonstrate that the incremental gains from targeting individuals within a household are twice—and sometimes several times—as large as those from targeting a household as a whole.

2. A Choice Model with Intrahousehold Heterogeneity

2.1. Why Intrahousehold Heterogeneity Matters

If purchase decisions are made by individuals, estimating parameters from data aggregated at the household level can yield parameters that are not only substantially different from the individual parameters but also different from any weighted average of the parameters (i.e., the household parameter need not fall in between the two). Without further information, there is no general relationship linking the individual-level estimated parameters and those from a household-level analysis. For example, consider the case where two price sensitive individuals in a household are equally likely

to make a shopping trip to buy from a category with two main brands. They have different preferences for brands, each one of them liking the brand the other one dislikes. They are unlikely to choose the brand they dislike, even if it is offered at a low price on a given shopping trip. Similarly, they will often purchase the brand they like, even if the price for it is relatively high in a given week. If we estimate a model knowing their individual identities, the individual-level choice parameters will reflect their strong preferences for the corresponding brands and an accurate price sensitivity. However, if we observed purchase patterns as a household (ignoring the individual identities), we would infer that this household sees the brands as interchangeable, sometimes buying a brand at a high price that was not purchased at a low price in a recent shopping trip, sometimes ignoring the low price of a brand even though it was purchased recently at the regular price. As a result, we would infer a price sensitivity that is *lower than that of either of the two individuals*; that is, we would not necessarily estimate a household price sensitivity that falls in between the price sensitivities of the individuals.

A manager sending targeted price discounts to price sensitive households might ignore this particular household based on such a household-level analysis. In the online appendix (Section A3), we provide details of a stylized example with two individuals purchasing two products according to a logit demand function and describe how the resulting aggregate choice pattern yields a low estimated price sensitivity parameter for different values of the parameter space.

In Section 2.2, we describe the problem of interdependent preferences and how that leads to correlated parameters in the utility function, and we develop a choice model that incorporates information on individuals that are linked by being members of the same household.

2.2. Correlated Choice Parameters

Since the seminal work by Davis (1976), we have known that the unit of decision making is neither the individual nor the monolithic household, but a system of decision makers that seek consensus or accommodate when making purchase decisions. Conjoint experiments have revealed that joint decisions can be described by combining individual preferences according to specific weights (Krishnamurthi 1988, Su et al. 2003). Previous surveys and studies on joint household decision making have described several possible decision strategies, such as averaging preferences, letting the person with stronger preferences decide (Aribarg et al. 2010), or engaging in complex bargaining (Corfman and Lehmann 1987, Davis 1976). Intentionally or not, each member of a household influences the purchases of the other household members (Simpson et al. 2012). More generally, the existing

literature suggests that individuals often take into consideration the preferences of other household members when they make a purchase. Thus, preferences of individuals within a household are likely to be interdependent, and a model of intrahousehold choice must account for the correlation in the individual-level purchase behavior within a household.

Much of the literature cited above relied on surveys or direct observation of individual consumption or preferences. We do not observe individual preferences or consumption behavior, only purchased choices. Therefore, we can develop a model that describes only these choices, but the available data do not allow us from making further inferences about the internal decision process or the consumption preferences of the individual household members. At the same time, we recognize that one household member's choices are not independent of the other household member's choices. For instance, a shopper in a household may choose a brand of yogurt because the other member likes it. However, given our data, we cannot disentangle the underlying mechanisms through which households choose grocery products following interdependent preferences. Consequently, we take a “reduced-form” approach and measure the revealed choice utilities of individuals from their sequence of choices while allowing for the utility parameters to be interrelated via a covariance matrix.

2.3. Brand Choice Intrahousehold Heterogeneity Model

We develop a hierarchical model to estimate the utility parameters for the household as a whole and for its individual members. The model assumes that the prior parameters for an individual are those of the household she belongs to. This hierarchical approach allows us to infer utility parameters from the individual choices and to simultaneously account for the interdependence of parameters within a household. The model partially pools individuals to their households, and the households to the population parameters. If we were to model individuals directly in a hierarchical model, those individuals with few observations would be shrunk to the population means.

We use a modified version of the traditional hierarchical logit model at the household level (Allenby and Rossi 1998, Rossi et al. 1996). Households choose among J brands or the outside good. We refer to individual i from household h as h_i , where $i \in \{1, 2\}$ for households with two members, $i \in \{1, 2, 3\}$ for households with three cards, and so forth. The utility of product j for individual i in household h on the household shopping trip t is

$$u_{hijt} = \alpha_{hij} + \beta_{hi} p_{jt} + \gamma_{hi} PROM_{jt} + \varepsilon_{hijt},$$

where α_{hij} is the individual h_i 's brand intercept parameter for brand j , β_{hi} is the price sensitivity parameters,

p_{jt} is the price of brand j on occasion t , γ_{hi} reflects individual h_i promotion sensitivity, $PROM_{jt}$ is the promotional activity of brand j at t (dummy variable with 1 indicating that there was a promotional sign), and the error term ε is the i.i.d. type I extreme value distributed component of utility that is observed by the household members but not by the researcher. Setting the utility of the outside good to zero yields the traditional logit probability of purchase as

$$P_{hijt} = \frac{\exp(\alpha_{hij} + \beta_{hi} p_{jt} + \gamma_{hi} PROM_{jt})}{1 + \sum_{k=1}^J \exp(\alpha_{hik} + \beta_{hi} p_{kt} + \gamma_{hi} PROM_{kt})}.$$

Introducing the notation $\theta_{hi} = (\alpha_{hij}, \beta_{hi}, \gamma_{hi})$ to designate the vector of parameters, we assume that $\theta_{hi} = \bar{\theta}_h + \eta_{hi}$, where $\bar{\theta}_h$ is a household-level parameter vector, and η_{hi} is an i.i.d. multivariate normally distributed deviation $\eta_{hi} \sim N(0, \Sigma_\eta)$ that measures how much intrahousehold variability there is in a given household. The diagonal elements of the covariance matrix Σ_η capture the variation among members of a household. The off-diagonal elements of Σ_η measure whether two parameters are likely to be positively or negatively correlated among household members. If all household members have similar preferences and typically agree on their purchases, they will have similar purchase patterns, and η_{hi} will be small and so will the elements of Σ_η . Obtaining the intrahousehold covariance matrix is an important advantage of this approach relative to estimating a traditional model using the individuals as a unit of heterogeneity. Admittedly, with enough data, this matrix could be computed ex post from the posterior distribution of the parameters. Yet our approach is likely to yield better results with the low number of individual observations that are typically available in grocery categories.

Let $\bar{\theta}_h = \bar{\theta} + v_h$ such that the household-level parameters are themselves multivariate normally distributed across households $\bar{\theta}_h \sim N(\bar{\theta}, \Sigma_\theta)$, with overall mean parameter $\bar{\theta}$ and covariance matrix Σ_θ . The unconditional distribution of the individual parameter is given by $\theta_{hi} \sim N(\bar{\theta}, \Sigma_\theta + \Sigma_\eta)$. In fact, for a household with two individuals, we can write the vector $\theta_{h\text{both}} = (\theta_{h1}, \theta_{h2})$ for both individuals, and its distribution conditional on $\bar{\theta}_h$ is

$$\theta_{h\text{both}} | \bar{\theta}_h \sim N\left(\begin{pmatrix} \bar{\theta}_h \\ \bar{\theta}_h \end{pmatrix}, \begin{bmatrix} \Sigma_\theta + \Sigma_\eta & \Sigma_\eta \\ \Sigma_\eta & \Sigma_\theta + \Sigma_\eta \end{bmatrix}\right). \quad (1)$$

We have demographic data at the household level. Demographic information is typically introduced in the model by assuming that the $\bar{\theta}_h$ depend linearly on demographic variables z_h and a multivariate normally distributed intercept according to $\bar{\theta}_h = \bar{\theta} + \Delta z_h + v_h$. Given our interest in comparing the magnitude of intrahousehold and interhousehold heterogeneity, and

for simplicity, we will estimate the model without demographic variables; that is, all variability will be captured by Σ_θ . The results using demographic variables are shown in the online appendix (Section A4.1) for purposes of comparison.

2.4. State Dependence

A direct way to introduce state dependence in a choice model (and the standard approach in the literature) is to add a dummy variable to the utility function that equals 1 if the corresponding brand was chosen in the previous purchase occasion (Keane 1997; Dube et al. 2008, 2010). We define the variable $I\{s_{ht} = j\}$, which equals 1 if the brand j was purchased on the household h 's previous purchase occasion. The parameter δ_h^1 is a household-level parameter that captures the influence that the previous household purchase has on the current purchase

$$u_{hijt} = \alpha_{hij} + \beta_{hi} p_{jt} + \gamma_{hi} \text{PROM}_{jt} + \delta_h^1 \cdot I\{s_{ht} = j\} + \varepsilon_{hijt}.$$

To extend this model to investigate the state dependence that may occur at the individual level when we observe choices of individuals in the same household, we need to consider that for a given shopping trip, we can define the previous brand chosen by the individual making the shopping trip as well as the previous brand chosen by the household, that is, by any individual in the most recent household shopping trip. We define the variable $I\{s_{hit} = j\}$, which equals 1 if the brand j was purchased on the individual h_i 's previous purchase occasion. If individual h_i was the one making the household's previous shopping trip, then $I\{s_{hit} = j\} = I\{s_{ht} = j\} = 1$. We define the following utility function:

$$u_{hijt} = \alpha_{hij} + \beta_{hi} p_{jt} + \gamma_{hi} \text{PROM}_{jt} + \delta_h^1 \cdot I\{s_{ht} = j\} + \delta_{hi}^2 \cdot I\{s_{hit} = j\} + \varepsilon_{hijt}. \quad (2)$$

In model (2), the parameter δ_h^1 plays the role of the traditional state-dependent effect, while δ_{hi}^2 measures the additional effect of the individual state dependence. Finally, we assume that all of the state dependence parameters are normally distributed, that is, $\delta_h^1 \sim N(\mu_{\delta^1}, \sigma_{\delta^1}^2)$ and $\delta_{hi}^2 \sim N(\mu_{\delta^2}, \sigma_{\delta^2}^2)$.

2.5. Estimation and Priors

The estimates are obtained by sampling the posterior distribution using Hamiltonian Monte Carlo (Duane et al. 1987) as implemented in the Stan language (Stan Development Team 2017).

The priors for the covariance matrices are based on the decompositions $\Sigma_\theta = \tau_\theta \Omega_\theta \tau_\theta$ and $\Sigma_\eta = \tau_\eta \Omega_\eta \tau_\eta$, where τ_θ and τ_η are matrices with standard deviations in the diagonals and zeros elsewhere, and Ω_θ and Ω_η are the corresponding correlation matrices. We assign Lewandowski et al. (2009) (hereafter, LKJ) priors to the correlation matrices Ω with a shape equal to 1.5. Larger

shape parameters bring the LKJ distribution closer to the identity matrix, and shape parameters between 1 and 2 are typically recommended to have flexible, relatively noninformative priors. The results are not very sensitive to higher parameter values, but the chains do not mix as efficiently. The priors for the standard deviations (including those in τ in the prior for the covariance matrix) are half-Cauchy with a scale parameter of 2.5 (Gelman 2006).

3. Data Description

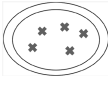

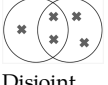

We use a data set from a major European retailer where we observe all of the purchases made during a six-month period (from May 2007 to November 2007) by a total of 19,626 households. Each household is identified by a loyalty card, which is presented to the cashier when they pay. Some multimember households have more than one loyalty card. (Rather than sharing one card for the whole household, the individual members use their own unique cards.) There are 3,440 households with two cards, 220 households with three cards, and 33 households with four or more cards.

Households with more than one card are demographically similar to those with more than one member and only one card. In terms of the number of household members, 16% of households have one member, 22% (22%) have two members, 23% (26%) have three members, and 26% (32%) have four members, where the numbers in parentheses are the percentages for households with more than one card. We observe approximately 5.1 million purchases in 494 categories spanning a broad range of products, from produce and meat to cleaning agents and toiletries. In addition to transactions, we have information on prices and on the products promoted in a given week.

We focus on a selected group of supermarket categories (coffee, pudding, bleach, dishwashing liquid, and kitchen rolls). As is common in the literature, we choose categories where a small set of brands account for a large proportion of sales in the category, products and packages are fairly standardized, and the purchase frequency is high enough to observe several purchases per household. This last requirement is important, as not all households have two loyalty cards, so we seek categories where we can observe purchases from both cardholders. We select households that have at least five purchases (different cutoff points yield similar results) in the category and where we observe at least one brand switch (this is necessary for identifying the state dependence parameters). A summary of the characteristics of each of these five data sets is shown in Table 1.

Households vary in how similar the brands chosen by the household members are. In some households, the individual members purchase the same set of brands (labeled "Identical"), while in others they

Table 1. Summary for All Five Data Sets

A. General description										
	Coffee		Pudding		Bleach		Dish wash. liq.		Kitchen roll	
Observations	17,455		17,190		5,986		5,145		18,152	
Households	1,568		1,475		533		517		1,592	
Households with multiple cards	213		184		67		61		275	
B. Classification of multicard households in terms of their brand choices										
	Coffee		Pudding		Bleach		Dish wash. liq.		Kitchen roll	
	HH (%)	Avg. purch.	HH (%)	Avg. purch.	HH (%)	Avg. purch.	HH (%)	Avg. purch.	HH (%)	Avg. purch.
Identical 	49	9.0	33	9.8	31	6.6	56	7.8	14	8.7
Included 	34	9.1	42	12.5	52	9.4	31	6.5	61	8.9
Some in common 	8	11.4	10	16.7	6	9.8	3	5.5	11	10.1
Disjoint 	10	7.1	14	4.8	10	7.7	10	7.3	14	6.8
C. Brand-level conditional shares and average prices										
	Coffee		Pudding		Bleach		Dish wash. liq.		Kitchen roll	
	Share	Avg. price	Share	Avg. price	Share	Avg. price	Share	Avg. price	Share	Avg. price
Brand 1	0.33	1.55	0.24	1.39	0.27	0.50	0.41	2.44	0.34	1.12
Brand 2	0.10	1.20	0.20	1.07	0.34	1.18	0.34	1.71	0.24	1.64
Brand 3	0.12	1.46	0.13	1.26	0.05	1.09	0.25	1.24	0.16	1.18
Brand 4	0.16	1.58	0.18	0.53	0.13	0.44			0.10	1.99
Brand 5	0.17	1.15	0.06	1.09	0.22	0.52			0.09	0.70
Brand 6	0.12	1.39	0.15	1.98					0.07	1.26
Brand 7			0.04	1.46						

Notes. Panel A provides a general overview of the number of observations, households (HHs), and households with multiple cards. Panel B classifies multicard households into groups. “Identical” refers to households whose members purchased the same set of brands. “Included” refers to those households where one or more members’ purchased brands are a subset of those of the other member(s). The label “Some in common” describes the case where the household members have some brands in common, as well as some brands that were only purchased by each individual. Finally, “Disjoint” refers to the case where household members purchased only brands that the other(s) did not. For each category, panel B shows the proportion of households falling into each case and the average number of purchases observed (“Avg. purch.”). Panel C shows information of market shares and average price for each brand in each category. Market shares are relative to the brands used in the estimation. They may not sum to one because of rounding. Average prices are in euros.

purchase completely different brands (“Disjoint”). We also have cases where one or more members purchase a subset of the brands of another member (“Included”). Finally, some households are composed of individuals who purchase some brand in common but each one of them also purchases brands that the others do not (“Some in Common”). The second panel of Table 1 shows the proportion of households with more than one card falling into each group for each data set, as well as the average number of purchases in each group. The third panel of Table 1 provides information on the average price and the market share (among those brands selected) for all five categories.

4. Results and Implications

In Sections 4.1 and 4.2, we describe the results from model (2) in the following order. First, we show model fit information that compares the proposed model to a model that uses disaggregate individual data but treats every individual as an independent shopper, and to a model that ignores the individual-level information (i.e., uses only household information). We then provide posterior summaries for all of the parameters for the coffee category. We later compare all five categories in terms of the brand choice state dependence and the variability of the utility parameters within and across households.

4.1. Model Comparison

We estimate two benchmark models to compare to our proposed model. In the first benchmark model, we estimate a traditional logit model where we estimate household-level parameters for the one-card households and individual-level parameters for the multicard households. In other words, we use all of the individual-level data but do not consider that individuals belong to a multicard household and are thus related to other individuals whose purchases are also part of our data set (we label this model the “individual model”). This individual model is a traditional logit model nested in our model. While our model considers that individual parameters for two individuals in a household h_1 and h_2 are distributed according to (1), the traditional individual model would simply assume

$$\theta_{h_{\text{both}}} | \bar{\theta} \sim N\left(\left(\bar{\theta}\right), \begin{bmatrix} \Sigma_{\theta} & 0 \\ 0 & \Sigma_{\theta} \end{bmatrix}\right).$$

This is the simplest model that would allow for individual-level targeting because it still yields an individual-level parameter estimate.

The second benchmark model (labeled the “household model”) considers only households and ignores the individual-level information; that is, the heterogeneity is modeled by estimating only a household-level parameter vector $\theta_{h_i} \sim N(\bar{\theta}, \Sigma_{\theta})$ and implicitly assuming $\theta_{h_{\text{both}}} = (\theta_{h_i}, \theta_{h_i})$.

All three models have exactly the same number of choices, and we compare them using the widely applicable information criterion (WAIC) (Watanabe 2010). Like the deviance information criterion, or DIC (Spiegelhalter et al. 2002), the WAIC uses existing data to approximate the expected predictive density for new data and penalizes model complexity, with the added advantage that it uses the whole posterior distribution of the parameters rather than a point estimate. It is invariant to parameterization and is asymptotically equal to cross validation (Gelman et al. 2014). Given the limited amount of data, we use in-sample validation for both WAIC and DIC.

The values² of the WAIC and DIC are lower for the proposed model than both the individual and the household models for each category, implying a better fit (Table 2). In addition, we compute the log Bayes factor (LBF) relative to the proposed model using the Gelfand and Dey (1994) approach. As a sampling distribution we use a multivariate normal distribution centered at the mean of the posterior draws and a covariance matrix proportional to the covariance matrix of the posterior draws. We tried different proportionality constants to “tune” the sampling distribution without finding a qualitative difference. The results in Table 2 use a tuning constant equal to 1.5. The negative values for the individual and household models favor our proposed model.

Table 2. Comparison Between Our Proposed Model and the Traditional Model Using Individual-Level Data

	Proposed model	Individual model	Household model
Coffee			
WAIC	31,161	31,402	31,253
DIC	32,384	40,191	40,299
OOS LL	−2,971	−2,978	−3,205
LBF	—	−48.8	−16.8
Pudding			
WAIC	36,404	36,752	36,444
DIC	36,537	46,343	47,378
OOS LL	−2,598	−2,655	−2,600
LBF	—	−1.1	−9.2
Bleach			
WAIC	12,052	12,437	12,475
DIC	12,414	15,124	15,125
OOS LL	−861	−873.1	−862
LBF	—	−7.0	−5.4
Dish wash. liq.			
WAIC	8,608	8,702	8,779
DIC	8,602	10,133	10,421
OOS LL	−616	−641	−701
LBF	—	−9.7	−1.4
Kitchen roll			
WAIC	43,756	44,033	44,028
DIC	43,696	55,366	55,357
OOS LL	−3,180	−3,190	−3,223
LBF	—	−15.5	−22.4

Notes. The WAIC and DIC are smaller for the proposed model, implying a better fit even penalizing for model complexity. The out-of-sample log likelihood (OOS LL) is higher in the proposed model than in both the individual and household models. Note that the OOS LL was computed using about 80% of the data for training and 20% for validation. The kitchen roll category was estimated without a promotion variable for the OOS LL because no promotions were available in the training set. Higher values of OOS LL imply a better fit. The LBF is the natural logarithm of the Bayes factor relative to the proposed model, using a Gelfand and Dey (1994) approach and a multivariate normal importance sampling distribution with the mean and covariance equal to that of the posterior draws. The values of the individual model and the household model are negative, favoring the proposed model.

We have six months of data, which we use in their entirety for this study to observe as many purchases as possible for each household and each individual. To check the out-of-sample prediction, we divide these six months into a five-month training data set and a one-month holdout sample. We estimate the model on the training data set. In Table 2, we provide the out-of-sample log likelihood values for each of the one-month holdout samples. It should be noted that in the case of the kitchen roll category, the training data did not contain any promotional activity, so the promotion variable was excluded for this comparison. On the whole, the evidence from the out-of-sample log likelihood favors the proposed model, but we recognize the limited duration of the data sets we are working with when interpreting these results.

The comparison just between the two benchmark traditional models (individual and household models) shows that there is no clear best approach. In some cases, the traditional model with individual-level information provides an improved fit because it can better accommodate heterogeneity in choice utility preferences via a larger number of parameters. In others, it seems that the model with individual-level information overfits the data and is penalized by its higher number of effective parameters or by providing a less accurate prediction in the holdout sample.

This model comparison analysis sheds light on what drives the improvement in performance of our proposed model: it strikes a balance of capturing individual-level variation and avoiding individual-level overfitting by shrinking individual-level estimates with few observations to the household-level estimates. Monte Carlo experiments validate these results (online appendix, Section A2).

4.2. Brand Intercepts and Price and Promotion Effects

The main results in this section and the targeting analysis in Section 5 are shown for the ground coffee category and only partially for the other categories.³ The coffee category is dominated by six brands (one of which is the store brand), accounting for more than 90% of category sales. In this particular market, ground coffee is sold in packages of 250 grams (about 9 oz.). The data set used for estimation has 1,568 households. For 213 of those households, we observe the purchases made by two or more individual members of the same household. The average number of purchases

per household is 11.1 (see Table 1, where we report the conditional sample shares).

The results are shown in Table 3. The first row shows the posterior means for the parameter $\bar{\theta}$, with the posterior standard deviations in parentheses. The parameter means for the brands are in agreement with the market shares and average prices for each brand. Even though the average effect of the promotions is not strong (i.e., close to zero), there is a large variability across households. The results show a significant sensitivity to prices.

As we showed earlier, the improved fit of the proposed model—particularly over the household model—comes from heterogeneity structure. However, the mean parameter values are similar. For comparison, Table 3 also shows the corresponding traditional model estimated at the household level (household model).

4.3. Comparing Inter- and Intrahousehold Variability

Let $\sigma_{\theta(kk)}$ and $\sigma_{\eta(kk)}$ be, respectively, the square roots of diagonal elements of the covariance matrices Σ_{θ} (across households) and Σ_{η} (within household). The parameter posterior means are shown in the second and third lines of Table 3. Our estimates show that heterogeneity parameters within a household for the brand intercepts are between 0.85 and 2.03. The intrahousehold heterogeneity parameters are smaller than the unobserved interhousehold heterogeneity parameters, meaning that—as expected—the variation in estimated brand intercepts across individuals within the same household is lower than that observed across households. The price and promotion heterogeneity coefficients show a different pattern. While still smaller

Table 3. Results for the Model with Intrahousehold Heterogeneity

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5	Brand 6	Price	Prom.	State dep. (HH) δ^1	State dep. (indiv.) δ^2
Model with intrahousehold heterogeneity (proposed model)										
Mean	2.38 (0.48)	−4.63 (0.56)	0.22 (0.49)	0.29 (0.51)	−1.04 (0.42)	−1.65 (0.53)	−2.47 (0.30)	0.09 (0.08)	0.74 (0.04)	2.17 (0.22)
$\sigma_{\theta(kk)}$ (interhousehold)	2.38 (0.26)	4.71 (0.36)	2.84 (0.33)	3.22 (0.44)	3.50 (0.24)	4.05 (0.32)	0.57 (0.22)	0.44 (0.11)	0.68 (0.06)	1.90 (0.18)
$\sigma_{\eta(kk)}$ (intrahousehold)	1.27 (0.39)	0.87 (0.43)	0.85 (0.31)	1.19 (0.43)	2.03 (0.34)	1.37 (0.35)	0.22 (0.12)	0.50 (0.30)		
Model without intrahousehold heterogeneity (household model)										
Mean	2.62 (0.45)	−4.59 (0.68)	0.45 (0.45)	0.56 (0.50)	−0.81 (0.39)	−1.49 (0.47)	−2.59 (0.29)	0.08 (0.07)	0.78 (0.03)	
$\sigma_{\theta(kk)}$ (interhousehold)	2.37 (0.24)	4.72 (0.55)	2.68 (0.22)	3.32 (0.25)	3.41 (0.21)	4.15 (0.42)	0.46 (0.24)	0.48 (0.11)	0.54 (0.05)	

Notes. This table shows the results for the proposed model and the household model (models with and without intrahousehold heterogeneity, respectively). The columns are the brand intercepts, the price coefficient, the promotion coefficient, and the two parameters driving the state dependence (δ^1 for the state dependence at the household (HH) level and δ^2 for that at the individual level). The row “Mean” refers to the posterior means of the brand intercepts and the price and promotion coefficients for each model. The rows labeled “ $\sigma_{\theta(kk)}$ ” and “ $\sigma_{\eta(kk)}$ ” show the square roots of the diagonal elements of the covariance matrices Σ_{θ} (interhousehold heterogeneity) and Σ_{η} (intrahousehold heterogeneity), as defined in Section 2.3. The standard deviations of the posterior are shown in parentheses.

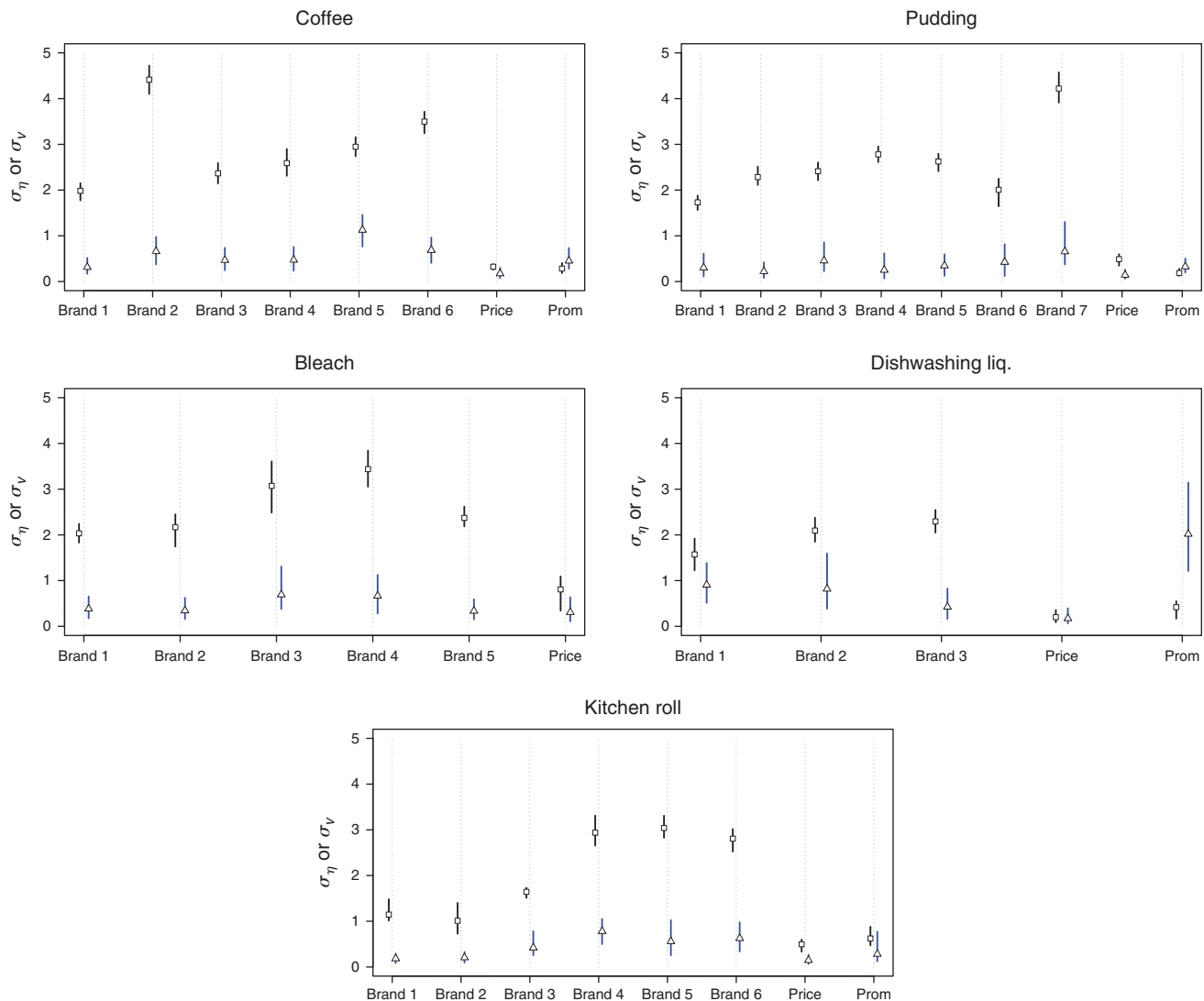
than the interhousehold heterogeneity, the intrahousehold heterogeneity parameters for the price coefficient are relatively large. In the case of the parameters for the promotion effect, they are similar in magnitude. This is an indication that the estimated brand utilities can be thought of as representing household preferences, while for the promotion sensitivities there is a strong individual component.

In Figure 1, we show the posterior estimates for $\sigma_{\theta(kk)}$ and $\sigma_{\eta(kk)}$ for all of the utility parameters and all of the categories analyzed. Each individual plot shows the mean posterior estimates with one standard deviation bar above and below the mean. In all but one brand in one category, the standard deviations of the heterogeneity of the brand intercepts across households are significantly larger than those for the heterogeneity

within households (the exception is dishwashing liquid, where the intervals for brand 1 overlap). The second pattern to note from Figure 1 is that the standard deviations for the price and promotion coefficients are typically similar in magnitude. In the case of the promotion coefficient, the intrahousehold heterogeneity is sometimes larger.

One convenient way to compare these two sources of heterogeneity is to compute the ratio of the unexplained variance among members of the same household to the unexplained variance among members of different households.⁴ Considering one parameter only (e.g., one brand intercept), the difference between two members of the same household is $\theta_{h_i} - \theta_{h_j} = \eta_{h_i} - \eta_{h_j}$, which has a variance of $2\sigma_{\eta(kk)}^2$. The variance between two individuals from different households is the variance of the

Figure 1. (Color online) Posterior Means for the Standard Deviations for the Inter- and Intra-household Heterogeneity



Notes. For each pair, the line on the left (square marker) represents σ_{θ} (interhousehold heterogeneity), and the line on the right (triangle marker) represents σ_{η} (intra-household heterogeneity). The vertical bars represent one standard deviation of the posterior distribution.

difference $\theta_{h_i} - \theta_{k_i} = \theta_h + \eta_{h_i} - (\theta_k + \eta_{k_i})$, which equals $2(\sigma_{\theta(kk)}^2 + \sigma_{\eta(kk)}^2)$. For a given category, we define the squared root of the ratio between these variances

$$\rho \equiv \sqrt{\frac{\sigma_{\eta(kk)}^2}{\sigma_{\theta(kk)}^2 + \sigma_{\eta(kk)}^2}},$$

as the metric for comparison. As this ratio approaches one (i.e., $\sigma_{\theta(kk)}$ is small), two individuals in a household are expected to be as similar as two individuals picked at random. To put this ratio into concrete terms, if $\sigma_{\theta(kk)} = \sigma_{\eta(kk)}$, then $\rho \approx 0.7$; if $\sigma_{\theta(kk)} = 2\sigma_{\eta(kk)}$, then $\rho \approx 0.45$; and values of ρ close to 0.2 imply that $\sigma_{\theta(kk)}$ is about five times larger than $\sigma_{\eta(kk)}$.

We compute this ratio for each draw from the posterior distribution and in Table 4 report the mean for the brand intercepts, the price coefficient, and the promotion coefficient for each of the categories we are working with. In the case of the brand intercepts, we show the average of this ratio rather than the individual ratios for each brand.

The ratios in Table 4 allow us to put the standard deviations on equal footing. Relative to the amount of interhousehold heterogeneity, the extent of intrahousehold heterogeneity in utility parameters is the least for the brand intercepts, followed by price sensitivity and then by promotion sensitivity. Indeed, in the case of promotions, two individuals in a household are likely to be as similar as two individuals picked from different households (with the exception of the kitchen roll category). Although we cannot make general statements for all grocery categories from our analysis of five categories, such a clear commonality across fairly diverse categories is indicative that there could be general patterns of utility parameters within a household.

4.4. Correlation of Utility Parameters

The estimated covariance matrices (shown together in Table 5) reveal a similar story. The upper triangle part of the matrix shows the interhousehold correlations from the heterogeneity matrix Σ_{θ} . These numbers should be interpreted as in the standard Bayesian

hierarchical models; for example, a household's intercept for the market-leader brand 1 negatively correlates with the intercept for the low-value brand 2 (−0.45), and with the price coefficient (−0.29).

The lower triangle of the matrix shows correlations for the intrahousehold heterogeneity Σ_{η} ; that is, the correlations capture the intrahousehold discrepancies. They are generally smaller than the correlations for the interhousehold heterogeneity; there is not much extra correlation beyond that already captured by the traditional heterogeneity. Note that they show a different pattern from the interhousehold heterogeneity. Consider, for instance, brands 1 and 4, which have a small correlation (−0.02) in the upper triangular matrix, that is, in the interhousehold heterogeneity. This means that a high household intercept for brand 4 does not provide any information about a high household intercept for brand 1. At the individual level, however, they are positively correlated (0.27), implying that, conditional on the household-level parameter, one individual's intercept for brand 1 often coincides with a higher intercept for brand 4 and vice versa.

4.5. State Dependence

In the last two columns of Table 3, we show the posterior summaries for the state dependence parameters for the coffee category. The parameter δ_h^1 is the purchase inertia dummy variable in the indirect utility function that equals 1 if the brand was chosen in the previous purchase by the household. The mean effect is positive (0.74) and within the same range as in previous studies of state dependence using similar models for grocery data (Seetharaman et al. 1999, Dube et al. 2008). There is also considerable heterogeneity, with most households showing positive state dependence (above 85% of households) and some negative. For the majority of households for which this parameter is positive, a choice of a particular brand in the previous shopping occasion by any individual increases the probability of that same brand being chosen again for all of the individuals within the household.

In the model with intrahousehold heterogeneity, we added an indicator variable intended to measure the effect of the previously purchased brand of the individual making the trip, with coefficient $\delta_{h_i}^2$. This parameter has a positive mean effect ($\mu_{\delta_2} = 2.17$). This result shows a strong effect of the individual state dependence: the contribution to the utility of having been the previous brand purchased by the individual shopper is higher than that of being the brand previously bought by the household but by a different shopper. In other words, the probability that an individual chooses a brand increases substantially if that brand was purchased by that individual on her previous shopping trip.

There is evidence of state dependence across all categories, as we would expect based on the existing literature. The estimates for μ_{δ_1} and μ_{δ_2} for all five categories

Table 4. Comparison of the Standard Deviations of Individuals Within Households and Two Individuals from Different Households

	Mean across brand intercepts	Price coefficient	Promotion coefficient
Coffee	0.21	0.50	0.86
Pudding	0.17	0.31	0.87
Bleach	0.20	0.42	—
Dishwashing liquid	0.38	0.71	0.98
Kitchen roll	0.22	0.32	0.50

Notes. The metric of comparison is ρ , defined as the square root of the ratio between the intrahousehold heterogeneity variance and the sum of the intra- and interhousehold heterogeneity variances. No promotions were observed for the bleach category.

Table 5. Correlations for the Utility Parameters for the Across-Household and Intra-household Heterogeneity

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5	Brand 6	Price	Prom.
Brand 1	1	−0.45	0.16	−0.02	−0.21	0.12	−0.29	0.09
Brand 2	−0.04	1	0.00	−0.23	−0.21	−0.65	0.22	0.03
Brand 3	−0.09	0.05	1	0.11	−0.04	0.06	−0.10	0.03
Brand 4	0.27	−0.07	0.01	1	−0.01	0.08	−0.46	0.38
Brand 5	−0.05	0.05	0.02	−0.25	1	0.03	0.05	−0.27
Brand 6	−0.10	0.05	0.13	−0.10	−0.11	1	−0.31	0.08
Price	−0.02	−0.05	−0.04	0.06	−0.23	−0.08	1	−0.21
Prom.	−0.05	−0.01	0.00	0.03	−0.20	0.08	0.06	1

Notes. The top-right numbers show the correlations from the covariance matrix for the interhousehold heterogeneity, Σ_θ . The bottom-left numbers show the correlations from the covariance matrix from the intrahousehold heterogeneity, Σ_η .

are shown in Table 6. Note that the posterior mean for μ_{δ_1} is positive, ranging from 0.33 to 0.92.

Inspection of the posterior summaries for the parameter μ_{δ_2} shows a very clear pattern across categories: The mean effect μ_{δ_2} is positive, and the relative size of μ_{δ_2} to the heterogeneity σ_{δ_2} shows that the value of $\delta_{h_i}^2$ is positive for the majority of households. Furthermore, $\mu_{\delta_2} > \mu_{\delta_1}$ for all categories. This implies a strong effect of the last brand purchased by the shopper, suggesting stronger average state dependence effects at the individual rather than the household level. This provides evidence that state dependence operates to a great extent on the individual decision maker compared to what the household purchased.

4.6. Robustness Checks

We ran a robustness check related to the category incidence specification. In our proposed model, we specify a brand choice model with a “no purchase” option.

Table 6. Posterior Estimates for the Mean and the Heterogeneity Standard Deviation of the Two State Dependence Parameters

	δ^1		δ^2	
	Mean	Heterog.	Mean	Heterog.
Coffee	0.74 (0.04)	0.68 (0.06)	2.17 (0.22)	1.90 (0.30)
Pudding	0.78 (0.04)	0.65 (0.06)	1.46 (0.16)	1.42 (0.18)
Bleach	0.33 (0.08)	0.69 (0.08)	1.29 (0.20)	0.42 (0.24)
Dishwashing liquid	0.92 (0.15)	1.48 (0.15)	3.49 (0.58)	2.65 (0.67)
Kitchen roll	0.64 (0.04)	0.54 (0.05)	1.11 (0.09)	0.66 (0.15)

Notes. The parameter δ^1 is the coefficient of an indicator in the utility function that equals 1 for the brand that was chosen in the previous shopping of any member of the household. The parameter δ^2 is the coefficient of an indicator in the utility function that equals 1 for the brand chosen in the previous shopping trip of the individual making the purchase. The numbers in parentheses are the standard deviations of the posterior means.

No purchase in the category is assumed to be the choice of the “outside” good. We estimated an alternative nested-logit model with category incidence at the upper level and brand choice at the lower level (Bucklin and Gupta 1992). The decision to purchase in the category conditional on a shopping trip increases with the attractiveness of the category, which in the context of a logit model is the natural logarithm of the denominator of the brand choice probability, namely, $CV_h = \ln[\sum_k \exp(\alpha_{hk} + \beta_h p_{kt} + \gamma_h PROM_{kt})]$. In this way, we model the probability of purchasing from the category as

$$\Pr(inc) = \frac{\exp(\iota_h + \iota_{2h} CV_h)}{1 + \exp(\iota_h + \iota_{2h} CV_h)},$$

where $\iota_h \sim N(\mu_{\iota}, \sigma_{\iota}^2)$ and $\iota_{2h} \sim N(\mu_{\iota_2}, \sigma_{\iota_2}^2)$ are heterogeneous parameters that govern the intrinsic category purchase propensity and sensitivity to the category value and vary across individuals.

The posterior estimates for this model are also similar to those of the model without the nested structure (see the online appendix, Section A4.3 and Table A5). Broadly speaking, this model uses ι_h to capture the choice for the outside good, rather than assuming that it is one of the choice alternatives. The estimation shows how our assumption of intrahousehold heterogeneity can be included in model extensions in a straightforward manner. However, given that there are no substantial changes by adding the category incidence model and that our focus is on comparing targeting policies with and without intrahousehold heterogeneity, we proceed in the remainder of this paper with the simpler assumption that the outside good is an extra option in the brand choice model, with the reassurance that, if the setting requires it, our model can be directly extended to accommodate other structural assumptions.⁵

As another robustness check, we estimated versions of the model using a category inventory variable (see the online appendix, Section A4.2 and Table A4). In addition, we used a control function approach to control for potential price endogeneity (Petrin and Train 2010, Rossi 2014). We used international coffee prices

and other categories whose supply fundamentals are correlated with coffee as instruments. The results and details are also shown in the online appendix (Section A4.4 and Table A6) and are not substantially different from the results presented here.

5. Implications for Targeting of Promotions

Household-level estimates of brand intercepts and price and promotional sensitivities can be used for targeting households with price discounts by customizing offers to specific households (Rossi et al. 1996, Zhang and Wedel 2009, Luo et al. 2013). To examine the value of the model with intrahousehold heterogeneity, we assume that a household (or an individual) can receive a targeted price discount in the form of a coupon, a message to their mobile devices, etc. While there are no coupons or targeted promotions in our data set, this assumption allows us to compare the potential gains of using individual-level parameters to target discounts tailored to the individual relative to targeting the household. In this analysis, we focus on the coffee category with some summaries and comments for the other categories.

5.1. Incremental Revenue for Individuals and for Households

We assume that a brand manager can send a targeted discount to either a household or an individual, who then purchases the product at a (nonstrategic) retailer. We assume a retail margin of $m = 35\%$, which is approximately the average retail margin of the retailer that produced the data, and we compute the expected incremental revenues for household h or individual h_i when presented with a proportional price reduction of $0 < F < 1$ for a product that sells at retail price p when not discounted. The manufacturer margin without the discount is $p(1 - m)$, and that with the discount is $p(1 - m - F)$. The probability of purchase of a discounted product j by household h can be written as $\Pr[j | \tilde{\theta}_h, s, \text{Price} = p(1 - F), \text{PROM} = 1]$, where PROM indicates that the item is on promotion. The incremental revenue is computed by averaging over the posterior draws $\tilde{\theta}_h$ for the individual unit of heterogeneity

$$R_h(F) = \frac{1}{n_{\text{draws}}} \sum_{\text{draws}} \Pr[j | \tilde{\theta}_h, p(1 - F), \text{PROM} = 1] \\ \times p(1 - m - F) - \Pr[j | \tilde{\theta}_h, p] \times p(1 - m).$$

Each household has a unique pattern for the incremental revenue (in euros). We can compute a similar incremental revenue value for the individuals in a household. Figure 2 shows the incremental revenue for three representative two-card households when offered a price reduction for brand 1. The nine charts in

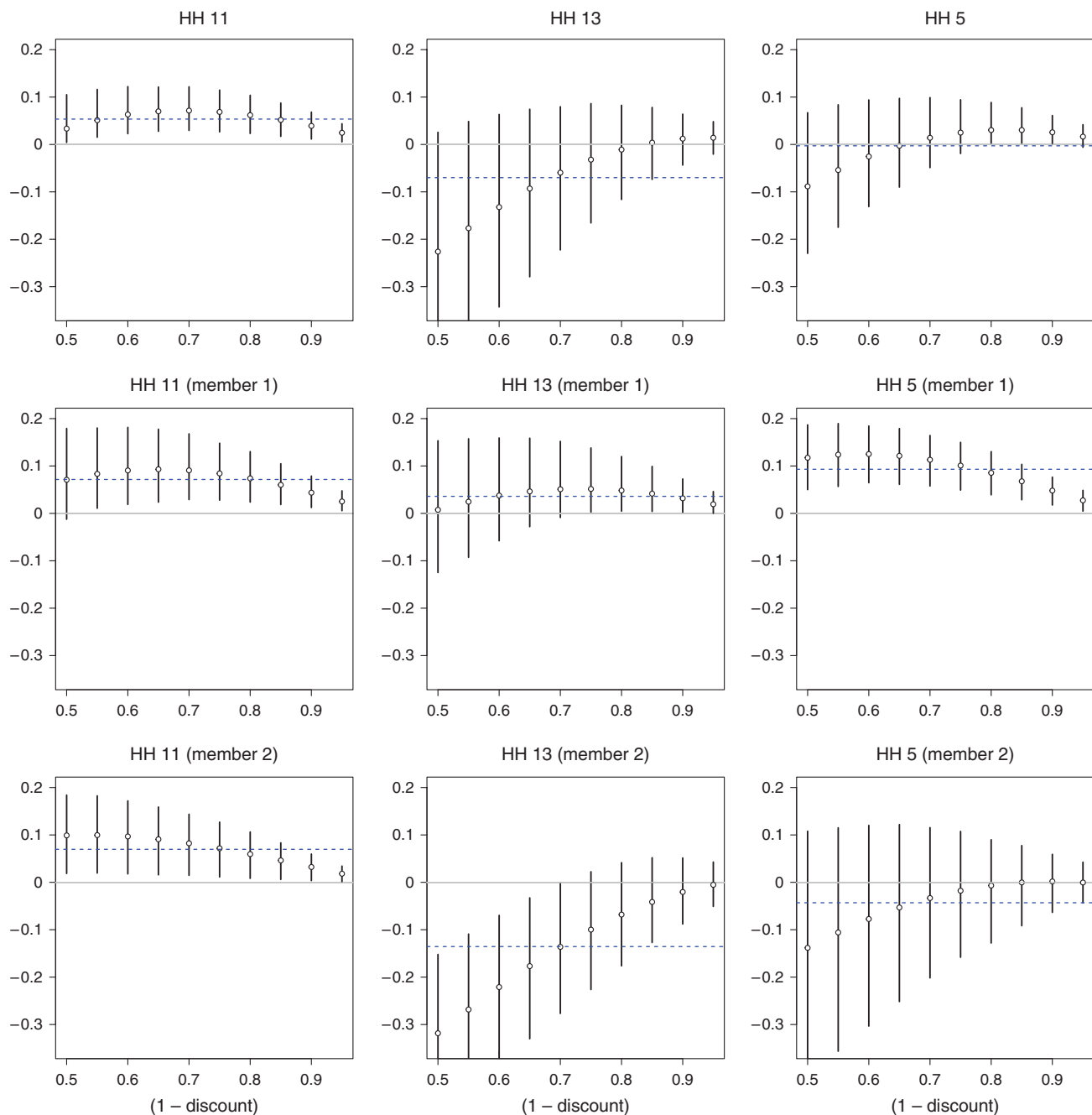
Figure 2 are arranged such that each “column” represents a household and each “row” represents either the household as an aggregate or an individual household member. The vertical lines and dot show the 20th–80th percentile and mean of the simulated incremental revenue for different discount levels, from 5% to 50% in 5% increments. The dotted line is the mean incremental revenue across all discount levels. Note that the x -axis is shown as $(1 - \text{discount})$.

The first column (HH 11) shows a household that is likely to yield a positive incremental revenue if offered a targeted price discount at any discount level. Higher discounts result in higher demand for the discounted products, but with diminishing returns because of the negative second derivative of the logit response function with respect to price. In other words, the expected revenue increases with demand up to the point where further discounting does not increase demand to compensate for the associated loss in unit margin. For the case of HH 11, the highest expected incremental revenue occurs at a discount of 30%. While bearing in mind that the discount levels are computed on a discrete grid, we refer to the discount level that yields the highest incremental revenue as “optimal.” The individuals in the household are also likely to produce a positive incremental revenue if they receive a discount targeted individually to them (first column, second and third rows). Note, however, that the optimal discounts for both individuals are deeper than that of the household, 35% and 45%, respectively. If these individuals receive a discount optimally designed for the household, they are still very likely to generate positive extra revenue, but the aggregate analysis would hide that the targeting firm is better off with a deeper discount, *for either of the individuals*.

The second column (HH 13) shows the case where the household as a whole does not seem like a good target for a price discount. The mean incremental revenue is negative, as is the expected incremental revenue at almost all discount levels. The exceptions are discounts between 1% and 15%, where the mean expected incremental revenue is positive but small, but still with a substantial probability of a negative outcome. This pattern hides the fact that one of the individuals (member 1, second column, second row) can yield a positive incremental revenue at around a 25% discount. In many practical applications, this household would not be targeted because of the poor expectations from the household analysis. Therefore, the targeting firm would miss the opportunity to generate incremental revenue from this household, and households like this.

The third column (HH 5) shows a more complex case, where individual-level analysis can add the most value. From the household chart (third column, first row) we see that a 20% discount has a positive expected

Figure 2. (Color online) Incremental Revenue for Three Representative Households



Notes. All of the results are for coffee brand 1 at different discount levels. In each chart, discount is shown in percentage discount for brand 1. The bars represent the 20% and 80% percentiles. The vertical axis represents the expected increase in revenue if the promotion targets the household (HH; first row) or household member 1 or 2 (second and third rows, respectively). The horizontal axis represents the retailer margin as a function of the promotional discount depth. The dashed line shows the mean across all discount levels.

incremental revenue. The problem is that a 20% discount will yield a much higher incremental revenue if it reaches member 1, and an expected loss in revenue if it reaches member 2 (second and third rows, respectively). The ideal discount would target member 1 with a 40% discount. By looking only at the household-level targeting analysis, managers risk reaching an

individual that is likely to generate a loss. If they reach the more profitable individual (member 2), they may do so at a suboptimal discount rate.

5.2. Comparing Promotion Policies

5.2.1. Shopper Model. Simply targeting each member of a household with the appropriate offer in place of

sending the household an offer based on data at the household level of aggregation would make the offer more *effective* if there is heterogeneity in the purchase behaviors of the household members. To improve the *efficiency* of the offer further would require us to know the household member likely making the shopping trip.

We estimate a reduced-form shopper model to predict the specific shopper from the household on each trip. We define a primary shopper by the number of purchases in the categories. The remaining individuals in the household (often just one other individual) are secondary shoppers. The decision of who shops is unlikely to depend on a specific category, but rather on the day of the week, who made the previous shopping trip, or how big or how long ago the past purchase was.⁶ Consistent with our model, conditional on who goes shopping, different categories and brands will have different probabilities of being purchased since the model parameters are individual specific in our specification. In that sense, who shops is indeed related to the purchase incidence and brand choice decisions, although, as we specify below, the expressions for the shopper model and the incidence and choice models are independent logit probabilities. If h_{it} is the primary shopper, we specify the probability of making a shopping trip relevant to the category (i.e., a purchase) as a logit model with covariates W

$$\Pr(h_{it} | \lambda_{h_i}) = [1 + \exp(-W_{h_{it}} \lambda_{h_i})]^{-1}.$$

For $W_{h_{it}}$, we include an intercept (parameter λ_{h1}) for household h that accounts for the baseline probability that the main shopper goes on the shopping trip, a variable that equals 1 if the main shopper did the previous

shopping trip to account for state dependence in shopping trips (with an associated parameter, λ_{h2}), and a dummy variable to indicate whether the purchase was made during a weekend (with an associated parameter, λ_{h3}). Finally, we add two variables that refer to the previous shopping trip with associated parameters λ_{h4} and λ_{h5} , respectively: the time since the last shopping trip of the household (in log days) and the number of items purchased in the last shopping trip (also logged). These two variables (time since last trip and number of items) are measured across categories, not necessarily the category we are investigating. We assume that the parameters of the shopper model have a normal distribution across individuals, that is, $\lambda_{h_i} \sim N(\bar{\lambda}_{h_i}, \Sigma_{\lambda_i})$, with a diagonal covariance.

Given a shopping trip made by a household (for whom the focal category being analyzed is relevant), we can predict the probability of the primary and secondary shoppers making the visit conditional on variables that are observed *prior* to the trip. The model is estimated for each category because the samples are different in each case.

The estimates (shown in Table 7) reveal mean intercepts (λ_{h1}) between 1.29 (dishwashing liquid) and 1.84 (kitchen roll), which are positive by design given our definition of the primary shopper. The value of the heterogeneity standard deviation (range, 0.22–0.49) implies that the probabilities of observing a purchase by the main shopper ranges between 83% and 89% for one standard deviation below and above the mean. The remaining variables also reveal heterogeneity in behavior. We estimated the state dependence in shopper choice. The parameter (λ_{h2}) in four out of the five categories is negative, indicating a tendency to alternate shopping trips. The evidence, however, is weak, as the posterior is not clearly on one side of zero.

Table 7. Posterior Estimates for the Mean and the Standard Deviation of Variables from the Shopper Model

Parameter	Coffee	Pudding	Bleach	Dishwashing liquid	Kitchen roll
Intercept (λ_{h1})					
Mean	1.71 (0.18)	1.36 (0.15)	1.77 (0.32)	1.29 (0.54)	1.84 (0.21)
Heterogeneity	0.26 (0.11)	0.49 (0.13)	0.42 (0.16)	0.27 (0.16)	0.22 (0.12)
State dep. (λ_{h2})					
Mean	−0.20 (0.18)	−0.38 (0.18)	0.15 (0.36)	−0.24 (0.39)	−0.31 (0.18)
Heterogeneity	0.79 (0.19)	0.83 (0.19)	0.64 (0.42)	0.79 (0.52)	0.87 (0.25)
Weekend (λ_{h3})					
Mean	0.04 (0.15)	−0.13 (0.18)	0.71 (0.45)	0.09 (0.38)	0.00 (0.16)
Heterogeneity	0.84 (0.17)	1.3 (0.25)	1.58 (0.57)	1.41 (0.62)	0.99 (0.21)
Time since last trip (λ_{h4})					
Mean	−0.15 (0.07)	−0.07 (0.06)	−0.01 (0.17)	0.07 (0.20)	0.01 (0.07)
Heterogeneity	0.10 (0.05)	0.10 (0.05)	0.21 (0.12)	0.11 (0.12)	0.06 (0.06)
No. of categories in last trip (λ_{h5})					
Mean	−0.04 (0.06)	0.10 (0.05)	−0.14 (0.12)	−0.03 (0.15)	−0.13 (0.06)
Heterogeneity	0.07 (0.03)	0.10 (0.06)	0.07 (0.05)	0.07 (0.05)	0.08 (0.03)

Notes. These parameters are the posterior estimates for a logistic model with the shopper as a dependent variable. Numbers in parentheses are the standard deviations of the estimates.

5.2.2. Aggregate Outcome of Targeted Promotions.

Taking all of the individuals or households into consideration, we can compute the average incremental revenue assuming different promotions policies. We keep the assumptions used above that either households or individuals can be targeted with a price promotion (coupon, short message service, etc.) at the household's or individual's optimal discount. We compare three types of promotional policies: a blanket promotion strategy (at 10% or 30% discount), a promotion targeted to the household that uses the optimal promotional depth computed from the estimated household-level parameters, and a promotion targeted to the individual who is likely to provide the highest incremental revenue, also at the optimal promotional depth according to the estimated individual-level parameters.

In computing the individual response to the promotion, we must also make assumptions on how the individuals within a household respond when receiving a discount targeted to the household. The simplest assumption is that each individual has an equal chance of using it ("50:50" in Table 8). Given that we have estimated a shopping incidence model, we can also assume that the probability of the individual using the promotion follows the probabilities computed in the shopper model ("shopper model" in Table 8). For simplicity, we set the variables to their means, and we use only the individual-level intercept as a source of heterogeneity. Finally, we can also assume that the promotion is used by the individual for whom it provides the highest level of utility according to our model ("highest utility" in Table 8). Of these three assumptions, the shopper model is the one that is closest to the actual behavior of the household. However, it can be argued that promotional usage is related to preferences, or obeys other unobserved processes, so it is instructive to compare the results under these different assumptions.

We compute the incremental revenue for a promotion sent, either to individuals or to the household. Note that if the incremental revenue for a particular brand under the assumptions of the promotional policy is negative at any promotional depth, the individual or household does not receive a promotion, and we set the incremental revenue to zero for that individual or household.

In Table 8, all promotional policies and household promotional behavior are listed with the corresponding average incremental revenue for brand 1 of the coffee category.

The blanket promotion yields practically zero or negative incremental revenues on average for both discount depths (10% and 30%). This implies that in our data set, a promotion for ground coffee does not generate extra expected revenue for a randomly chosen customer. The targeted promotion yields a higher incremental revenue by offering promotions only to

Table 8. Results Based on Different Promotion Strategies for One Brand

Promotion policy	Who shops?	Incremental revenue per targeted promotion brand 1 (euros)
10% blanket	50:50	0.000
10% blanket	Shopper model	0.001
10% blanket	Highest utility	0.003
30% blanket	50:50	−0.012
30% blanket	Shopper model	−0.009
30% blanket	Highest utility	−0.001
Optimal to the HH	50:50	0.033
Optimal to the HH (A)	Shopper model	0.035
Optimal to the HH	Highest utility	0.032
Optimal to individual (B)		0.052
Improvement (ratio (B − A) / A)		48%

Notes. All results are from the coffee category, and the analysis is done for brand 1. "Blanket" refers to the same targeted discount sent to all households. "Optimal to the HH" means that each household received a discount according to the maximum expected revenue for the household. "Optimal to individual" means that the individual with the highest expected incremental revenue received a targeted price discount at the optimal discount level.

those households or individuals who will use it to buy the brand, a similar result as that reported by Rossi et al. (1996). When we consider the promotion targeted at the individual level, the incremental revenue increases by 48% relative to targeting households. We have shown the price in monetary terms to facilitate the comparison to the literature.

To the extent that promotional activity is important in our setting, individual-level targeting offers improvement in performance. In Table 9, we show the average improvement across all brands and a selection of relevant comparisons for the three categories where promotions are observed or relevant, namely, coffee, dishwashing liquid, and kitchen roll. (There are no promotions in the bleach category, and they are too small in the pudding category for the analysis to be worthwhile.) These values are also positive and in line with what we would expect from adding targeting information to a promotional campaign. Table 9 also shows the improvement relative to the 10% blanket promotion. Note that these improvements are sometimes higher than the incremental revenue. For instance, in the case of dishwashing liquid, the incremental revenue is 0.8% of price, but there is an improvement of 3.8% over the blanket 10% discount. The average incremental revenue for some brands of using a blanket 10% discount is sometimes negative because the loss of margin is not made up by the increased choice probability from a lower price. More importantly, there is a positive

Table 9. Results Based on Different Price Promotion Strategies Under Different Assumptions of How the Promotion Is Used

	Coffee		Dishwashing liquid		Kitchen roll	
	Euros	Relative to price (%)	Euros	Relative to price (%)	Euros	Relative to price (%)
Average incremental revenue across brands for promotion targeted to the individual	0.028	2.6	0.011	0.8	0.002	0.2
Improvement over 10% blanket	0.022	2.1	0.051	3.8	0.010	1.0
Improvement over the promotion optimally targeted to the household	0.008	0.9	0.010	0.5	0.003	0.2
Results for the brand with best results						
Average incremental revenue for promotion targeted to the individual	0.056	5.3	0.023	1.7	0.004	0.4
Ratio						
A. Ratio of incremental revenues of targeting the individual to targeting the household		1.5		7.3		12.6
B. Ratio of incremental revenues of targeting the individual to targeting the household, when taking the next shopping trip into account		2.1		8.5		17.3

Notes. The top three lines replicate the results in Table 8, but for all brands. The bottom three lines show the results for the brand with the best results. Further improvement of line B over line A comes from simultaneously considering the expected incremental revenue of the current shopping trip and the probability of making the next shopping trip.

increase in incremental revenue by targeting individuals rather than households.

We also show the results for the brand with the highest increases in revenue, as we do not expect brands with small increases or negative changes to revenue to run promotions of this kind. We compute the ratio of the incremental revenue of targeting the individual to that of targeting the household. In the case of coffee, for instance, one brand shows a 5.3% increase in revenues. If we compare this result to that of targeting the household by computing the ratio of the corresponding incremental revenues, we find that individual targeting yields 1.5 times more revenue (Table 9, row A). This is the same result as shown in the last line of Table 8 (as a ratio instead of a percentage improvement, that is, 48% increase is 1.5 when computed as a ratio), and it assumes that every promotional message is actually considered (not necessarily used, as households and individuals could always choose the outside good). Note also that this ratio is 7.3 and 12.6 for the dishwashing liquid and kitchen roll categories, respectively.

The shopper model can be used to fine-tune the targeting of the promotion. For some households, there could be two factors on deciding whom to target. The company should target those individuals whose incremental revenue is largest conditional on making a shopping trip. In addition, the company should consider the probability of an individual making a shopping trip next. In some cases, it could be optimal to choose the individual with lower expected incremental revenue because that individual is more likely to make the shopping trip. Row B in Table 9, computes the

improvement ratio under this additional assumption: that a promotion, even if desirable for the individual or household and profitable to the firm, may not be used if the individual is not the next one to make the shopping trip. The individual-level promotion in this case considers whom to target in the household based not only on profitability, but also on probability of making the next trip. The improvement is noticeably higher: 2.1 times for coffee, 8.5 times for dishwashing liquid, and 17.3 times for kitchen rolls. While these large multiples show a clear improvement, it should be pointed out that they also reflect the small incremental revenue of a promotion without the individual-level targeting.

These results suggest that the potential improvement for individually targeted price discounts is driven by three important factors. First, they allow managers to choose the individual within the household with the highest expected incremental revenue. This individual, as we have seen, may belong to a household for which the computed expected revenue is not positive, if considered as a whole. Second, the targeted message can offer a discount level that is optimal to that individual. We showed above that the discount level that is optimal at the individual level does not necessarily correspond to that of the household. Finally, as shown in Table 9, the targeted discount should consider the probability of an individual making a shopping trip; that is, even if one individual in a household shows the highest expected revenue conditional on making a shopping trip, a second household member may yield a higher expected revenue once we consider the probability of making the shopping trip.

6. Conclusion

This paper studies a potentially important aspect of the heterogeneity of demand: the variation in purchasing behaviors *within* a household. It is the first study to propose a model to infer the utility parameters of individuals within a household from scanner data (rather than surveys of experiments).

The choice model we present captures the intrahousehold variability in brand intercepts and sensitivities while allowing for the interdependence in household members' choices that has been reported in the literature for the last four decades. While the specific mechanism underlying this interdependence remains unobserved, we show that our model can be derived from one in which each household member takes the other's preferences into account while making her choices. Empirically, the model parsimoniously captures the joint decision making revealed in the data through a common prior and a multivariate deviation from that prior according to an empirically estimated covariance structure. Our model is particularly useful in situations where using household-level information compensates for the lack of information on one of the decision makers.

Our analysis also sheds light on the concept of loyalty. While loyalty is traditionally defined at the level of the individual decision maker, defining who the decision maker is depends on the level of aggregation that we observe. We have shown that both household-level loyalty and individual-level loyalty matter. If the disaggregated data are available to analysts and managers, they should consider both levels of aggregation to paint a more complete picture of the household. We apply the model to simulate the profitability of price discounts. We show that using individual-level utility parameters and price promotions targeting household shopping behavior can substantially increase the profitability of the promotions.

Most of our results indicate that there are commonalities across categories. The individual state dependence is clearly stronger than the household state dependence, in some cases several times larger. In addition, the relative importance of the intrahousehold heterogeneity follows a very similar pattern across categories. When considering brand intercepts, the standard deviation intrahousehold is about a fourth of the standard deviation across households. In terms of promotions, two individuals living in a household are expected to be as similar as two individuals picked at random from the population. In our data, we had limited overlap of households with multiple cards and purchasing in several categories. This limitation prevented us from developing a multicategory analysis in the spirit of Ainslie and Rossi (1998).

There are still questions that cannot be addressed because of limitations of our current data. For instance,

it would be beneficial for targeting decisions to observe individual rather than household demographics. We have remained agnostic regarding the process that household members follow when making purchase decisions and when they consume the products they purchase. In contexts where consumption information is available, our model could be extended to provide a fuller picture of the joint behavior of household members. For instance, if we are modeling the consumption of music on mobile devices, we could make stronger assumptions about individual consumption preferences, not just choice utilities. Furthermore, observing individual consumption and joint consumption of products and services (e.g., video streaming on an individual platform such as a smartphone, or a potentially shared platform such as a TV) could help us identify not only individual preferences but also some aspects of the joint decision process.

The problem of joint decision making can only become more important as these kinds of data sets become more common. Most loyalty card programs allow multiple cards for a household, and the multiple cards can be differentiated by the provider of the program (e.g., Tesco Club Card). Thus, data may be collected and used for testing, prediction, and optimization. Airline loyalty programs sometimes know the family connections of their members. More recently, services offered by Apple iTunes know which accounts are connected and can make predictions and recommendations about one of its members by looking at purchases from the other member.

Acknowledgments

The authors would like to thank seminar participants at the Judge Business School and participants at the 2017 German Marketing Commission Symposium for helpful comments and suggestions. The authors would also like to extend our gratitude to three anonymous reviewers, the associate editor, and the senior editor for insights and constructive comments that substantially improved this paper.

Endnotes

¹This means that the variability in the parameters of individuals picked at random from the population can be described by a combination of our estimated inter- and intrahousehold heterogeneity.

²We compute the WAIC using the parameter correction p_{waic2} rather than p_{waic1} , as it is recommended in the cited literature. The results using p_{waic1} follow the same pattern in all cases. The computation of the WAIC was coded in R following Gelman et al. (2014, Equations (12) and (13), p. 1003).

³This is due to space considerations. However, we will discuss broad patterns across categories. Detailed results are available from the authors on request.

⁴We thank an anonymous reviewer for suggesting the idea behind this comparison of the variances.

⁵Detailed results are available from the authors on request.

⁶We recognize that both cardholders may be physically present during a trip, but since only one loyalty card is used on a given trip, we focus on that individual as being the person making the trip.

References

- Ainslie A, Rossi PE (1998) Similarities in choice behavior across product categories. *Marketing Sci.* 17(2):91–106.
- Allenby GM, Rossi PE (1998) Marketing models of consumer heterogeneity. *J. Econometrics* 89(1):57–78.
- Aribarg A, Arora N, Kang MY (2010) Predicting joint choice using individual data. *Marketing Sci.* 29(1):139–157.
- Bhat CR, Pendyala RM (2005) Modeling intra-household interactions and group decision-making. *Transportation* 32(5):443–448.
- Browning M, Chiappori PA (1998) Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica* 66(6):1241–1278.
- Bucklin RE, Gupta S (1992) Brand choice, purchase incidence, and segmentation: An integrated approach. *J. Marketing Res.* 29(2):201–215.
- Corfman KP, Lehmann DR (1987) Models of cooperative group decision-making and relative influence: An experimental investigation of family purchase decisions. *J. Consumer Res.* 14(1):1–13.
- Davis HL (1976) Decision making within the household. *J. Consumer Res.* 2(4):241–260.
- Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. *Phys. Lett. B* 195(2):216–222.
- Dube JP, Hitsch GJ, Rossi PE (2010) State dependence and alternative explanations of consumer inertia. *RAND J. Econom.* 41(3):417–445.
- Dube JP, Hitsch GJ, Rossi PE, Vitorino MA (2008) Category pricing with state-dependent utility. *Marketing Sci.* 27(3):417–429.
- Gelfand AE, Dey DK (1994) Bayesian model choice: Asymptotics and exact calculations. *J. Royal Statist. Soc. Series B (Methodological)* 56(3):501–514.
- Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* 1(3):515–534.
- Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Statist. Comput.* 24(6):997–1016.
- Gupta S, Steckel JH (1993) Preference aggregation and repeat buying in households. *Marketing Lett.* 4(4):321–336.
- Keane MP (1997) Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* 15(3):310–327.
- Krishnamurthi L (1988) Conjoint models of family decision making. *Internat. J. Res. Marketing* 5(3):185–198.
- Ladner P (2009) Spousal control and intra-household decision making: An experimental study in the Philippines. *Amer. Econom. Rev.* 99(4):1245–1277.
- Lewandowski D, Kurowicka D, Joe H (2009) Generating random correlation matrices on vines and extended onion method. *J. Multivariate Anal.* 100(9):1989–2001.
- Luo X, Andrews M, Fang Z, Phang CW (2013) Mobile targeting. *Management Sci.* 60(7):1738–1756.
- Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* 47(1):3–13.
- Rick SI, Small DA, Finkel EJ (2011) Fatal (fiscal) attraction: Spendthrifts and tightwads in marriage. *J. Marketing Res.* 48(2):228–237.
- Rossi PE (2014) Invited paper—Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Sci.* 33(5):655–672.
- Rossi PE, Allenby GM (1993) A Bayesian approach to estimating household parameters. *J. Marketing Res.* 30(2):171–182.
- Rossi PE, Allenby GM, McCulloch R (2012) *Bayesian Statistics and Marketing* (John Wiley & Sons, Hoboken, NJ).
- Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–334.
- Seetharaman PB, Ainslie A, Chintagunta PK (1999) Investigating household state dependence effects across categories. *J. Marketing Res.* 36(4):488–500.
- Simpson JA, Griskevicius V, Rothman A (2012) Consumer decisions in relationships. *J. Consumer Psych.* 22(3):304–314.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *J. Royal Statist. Soc. Series B (Methodological)*. 64(4):583–639.
- Stan Development Team (2017) Stan Modeling Language: User's Guide and Reference Manual, Version 2.16.0. <http://mc-stan.org>.
- Su C, Fern EF, Ye K (2003) A temporal dynamic model of spousal family purchase-decision behavior. *J. Marketing Res.* 40(3):268–281.
- Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Machine Learn. Res.* 11(December):3571–3594.
- Zhang J, Wedel M (2009) The effectiveness of customized promotions in online and offline stores. *J. Marketing Res.* 46(2):190–206.