

Who's got the coupon? Estimating Consumer Preferences
and Coupon Usage from Aggregate Information

Andrés Musalem

Eric T. Bradlow

Jagmohan S. Raju*

October 2007

*Andrés Musalem is Assistant Professor of Marketing at The Fuqua School of Business of Duke University. Eric T. Bradlow is The K.P. Chao Professor, Professor of Marketing, Statistics and Education and Academic Director of the Wharton Small Business Development Center, The Wharton School of the University of Pennsylvania. Jagmohan S. Raju is The Joseph J. Aresty Professor, Professor of Marketing and Executive Director of the Wharton-Indian School of Business Program, The Wharton School of the University of Pennsylvania. The authors thank participants at the XXVII Marketing Science Conference at Emory University and seminars at Cornell University, Dartmouth College, Duke University, Harvard University, INSEAD, Purdue University, Tilburg University, University of Maryland, University of North Carolina, University of Pennsylvania (Wharton), University of Rochester, University of Southern California, University of Texas at Austin, University of Texas at Dallas and Washington University in St. Louis. In addition, David Schmittlein, Ed George and David Bell provided excellent comments. We also thank the editor Russell S. Winer and two anonymous reviewers for their helpful suggestions that have greatly improved this paper. Please address all correspondence on this manuscript to Andrés Musalem, 1 Towerview Drive, Durham, NC 27708, Phone: (919) 660-7827, Fax: (919) 681-6245, amusalem@duke.edu.

Who's got the coupon? Estimating Consumer Preferences and Coupon Usage from Aggregate Information

Abstract

Most researchers in the Marketing literature have typically relied on disaggregate data (e.g., consumer panel) to estimate the behavioral and managerial implications of coupon promotions. In this article, we propose the use of individual-level Bayesian methods for studying this problem when only aggregate data on consumer choices (market share) and coupon usage (number of distributed coupons and/or number of redeemed coupons) are available. The methodology is based on augmenting the aggregate data with unobserved (simulated) sequences of choices and coupon usage consistent with the aggregate data. Different marketing scenarios are analyzed, which differ in terms of their assumptions about consumer choices, coupon availability and coupon redemption.

Initially, we consider a situation where the researcher observes aggregate market shares, marketing activity, the number of distributed coupons redeemed and the number of coupon redemptions for each brand in each period. Then, we generalize the estimation procedure to handle more realistic situations. These generalizations include: i) the researcher observes the number of redeemed coupons in each period, but not the total number of consumers that received a coupon, ii) consumers use coupons only when its redemption enhances the utility of the chosen alternative and iii) firms may coordinate their coupon distribution policy with other elements of the marketing mix.

The proposed methods are illustrated using both simulated data and a real data set for which an extensive set of posterior predictive checks are used to validate the aggregate-level estimation. In addition, we also relate our empirical results to some of the findings in the literature about the coordination of coupon promotions and pricing and we show how our methodology can be used to answer relevant managerial questions normally reserved for panel data, such as the analysis and comparison of alternative coupon targeting policies.

Key Words: *Coupon promotions, random-coefficients choice models, structural demand models, data augmentation, Markov chain Monte Carlo.*

1 Introduction

Consumer packaged goods (CPG) manufacturers invest billions of dollars every year in coupon promotions. According to CMS, Inc., 323 billion coupons were issued in 2005 with an average face value of \$1.16, from which only 3 billion coupons were redeemed. Determining the impact of coupon promotions has intrigued academics and practitioners alike for decades. One important input into this process is a clearer understanding of the preferences and characteristics of those who redeem coupons and those who do not.

When individual-level data specifying the choices and coupon usage of each panelist in each period are available, it becomes easier to study these issues. This is the case in most of the recent studies in the Marketing literature where *primary* data have been collected using surveys or field experiments (e.g., Krishna and Shoemaker 1992; Bawa, Srinivasan and Srivastava 1997). Similarly, in the case of *secondary* data, researchers have often used disaggregate information (consumer panel data) to estimate the behavioral and managerial implications of coupon promotions (e.g., Narasimhan 1984; Bawa and Shoemaker 1987; Neslin 1990; Chiang 1995; Leone and Srinivasan 1996; Erdem, Keane and Sun 1999), or made these inferences from the estimated consumer price sensitivities (Rossi, McCulloch and Allenby 1996). However, when this is not “directly possible” because only aggregate information is available (such as market shares and number of redeemed coupons), researchers have typically used simplified approaches. For example, some researchers have used demand models that are not explicitly linked to individual-level assumptions of utility maximization (e.g., Nevo and Wolfram 2002). Other researchers have only focused on coupon redemption without explicitly modelling brand choice (e.g., Reibstein and Traver 1982; Lenk 1992). Similarly, reduced-form models have been used to study the efficiency of coupon promotions (e.g., Anderson and Song 2004), while some researchers have estimated the impact of coupons by treating them as price reductions (e.g., Besanko, Dubé and Gupta 2003).

Recent advances in Bayesian analysis of aggregate data initially proposed by Chen and Yang (2006) and extended in Musalem, Bradlow and Raju (2006), provide new tools for

the estimation of demand models that are formulated as the aggregation of individual-level choice models. Using a generalization of these Bayesian techniques, a new methodology is presented here, which is based on augmenting the observed aggregate data (e.g., market shares, number of redeemed coupons) with unobserved (simulated) sequences of choices *and* coupon usage. As a result, the proposed methodology exhibits all the benefits of Bayesian methods (e.g., no asymptotic assumptions are needed and prior information can be easily incorporated). In addition, the model is built from micro-level assumptions (e.g., utility maximization), which enables a researcher not only to justify the proposed model using marketing or economic theories, but also to estimate the distribution of preferences among consumers and, consequently, the impact of coupon promotions on the patterns of individual choices and coupon usage using only aggregate data.

In terms of the contribution of this paper over existing methodologies, we note that both Chen and Yang (2006) and Musalem, Bradlow and Raju (2006), assume that the independent variables (e.g., prices and promotion) are constant across consumers. We note that this is also true in most papers in the marketing and economics literature that rely on aggregate data (e.g., Berry 1994, Romeo 2007). In the case of coupon promotions, this assumption is certainly not appropriate, as not all consumers have access to or are willing to redeem coupons. Consequently, and from a broader perspective (i.e., beyond this specific application to coupon promotions), a major contribution of this research is the development of a methodology that allows a researcher to estimate a demand model with aggregate information on both dependent and independent variables. In addition, coupons are only observed when coupon redemption takes place. This implies that the information on this independent variable (coupon availability) is not only aggregate, but also censored. Therefore, the proposed methodology enables a researcher to solve not only this aggregation problem on both dependent and independent variables, but also to resolve these censoring issues. This is accomplished by inferring coupon availability from coupon redemption data as in Erdem, Keane and Sun (1999), although in this paper this is achieved using only aggregate

information.

In terms of the demand model under study, several scenarios are analyzed, which differ on their assumptions about consumer behavior and available information. In the simplest case, we consider a situation where the researcher observes the aggregate market share, marketing activity, number of coupons redeemed and number of consumers holding a coupon for each brand in each period. The estimation procedure is then generalized in order to handle more realistic situations regarding data availability and to incorporate assumptions consistent with utility maximization and the strategic behavior of firms. Specifically, these generalizations include:

1. Number of coupons distributed is unknown: the researcher knows the number of redeemed coupons in each period, but not the total number of consumers that received a coupon.
2. Utility Maximization: a coupon is used only if the utility associated with its redemption is non-negative.
3. Strategic Coordination of Marketing Activity: firms may strategically align their coupon distribution with other elements of the marketing mix (e.g., prices, feature and display).

The proposed methods allow a researcher to answer important managerial questions such as determining the penetration of coupons (i.e., the fraction of consumers that have used a coupon at least once), the number of heavy users of coupons (i.e., the fraction of consumers that have used at least k coupons) and the expected number of users that would switch from one brand to another if they received a coupon; again all of this from aggregate data. For each of these quantities of interest, it is possible to not only compute a point estimate, but also estimate its entire posterior distribution using Markov chain Monte Carlo (MCMC) simulation, the computational approach utilized here. This is an important advantage of the proposed method in comparison to reduced-form approaches as understanding the variability in these quantities can also play a role in decision making.

In addition, as we will show in Section 6, this methodology allows a researcher to simulate the effects of policy experiments, such as estimating the performance of alternative coupon targeting strategies (e.g., targeting all category buyers versus targeting only those who bought a particular brand). Moreover, the aggregate estimation results can also be used to estimate the value of purchasing or collecting individual data for targeting purposes.

In summary, the main contribution of this research is the development of a new methodology to measure the impact of coupon promotions on consumer choice using only aggregate data on both choices (dependent variable) and coupons (independent variable) and overcoming coupon availability censoring problems. Using this methodology it is possible to account for heterogeneity in consumer preferences, answer relevant managerial questions, and analyze the consequences of different couponing strategies. Moreover, the methods presented here may also become useful to researchers dealing with aggregate data on both dependent and independent variables in other marketing and economics settings. Therefore, we hope that this paper will contribute in expanding the set of tools that researchers may consider when dealing with aggregate data.

The rest of this paper is organized as follows. In Section 2, a simple case is described which illustrates the basic ideas of the proposed methodology and introduces the needed notation. In Section 3, the estimation procedure is generalized by assuming that the researcher has data on the number of redeemed coupons, but *not* about the total number of consumers that received a coupon. In Section 4, a utility maximization framework is introduced for the coupon usage decision under which coupons are redeemed only if the utility associated with their redemption is non-negative. Furthermore, this section also considers the possibility that manufacturers might strategically align their coupon distribution with other marketing variables (e.g., prices, feature and display), and that prices might be correlated with unobserved market conditions (price endogeneity), where these unobserved demand shocks may exhibit autocorrelation. In Section 5, a real data set of purchases is analyzed for which the distribution of consumer preferences is estimated with and without knowledge of the

individual purchases and coupon usage. In Section 6, it is shown how these results can be used to analyze important managerial questions related to the estimation of the value of information and the evaluation of alternative coupon distribution approaches. Finally, Section 7 concludes this paper with a discussion of interesting avenues for future research.

2 The Basic Model

Assume N consumers make purchase decisions in each of T periods choosing among J brands in a given market. In each period t , $N_{jt}^c \leq N$ consumers receive a coupon for brand j and $N_{jt}^r \leq N_{jt}^c$ of them redeem their coupons. In addition, we specify a set of initial assumptions for this basic model regarding information availability and coupon distribution and redemption. In particular, the following assumptions are specified:

- (A1) **Multiple Coupons.** Each consumer may have coupons for more than one brand in each period.
- (A2) **Available Information: Distributed Coupons and Redeemed Coupons.** The researcher observes aggregate data regarding the total number of distributed coupons and the total number of coupon redemptions for each brand in each period.
- (A3) **Immediate Expiration.** Coupons are only valid for one period.
- (A4) **Coupon Distribution.** Each consumer has the same probability of being among the N_{jt}^c consumers that received a coupon (i.e., no targeting by manufacturers).
- (A5) **Coupon Redemption.** If a consumer has a coupon for brand j and chooses to buy brand j in a given period, then she redeems her coupon in that period (we note that this assumption will be generalized in Section 4).

We note that some of these assumptions are only used for simplifying purposes (e.g., **A5**), however, in the next sections, we will introduce modifications to the basic model in order

to extend the proposed methodology to more general settings. Specifically, modifications associated with **A2**, **A3**, **A4** and **A5** will be described; one is not needed for **A4** as it is already general.

In addition to this set of assumptions about coupon usage, it is assumed that consumers choose the product with the highest utility and that the choice of consumer i in period t (y_{it}) satisfies:

$$\begin{aligned}
 (1) \quad y_{it} &= \operatorname{argmax}_j U_{ijt} \\
 &= \operatorname{argmax}_j V_{ijt} + \epsilon_{ijt} \\
 &= \operatorname{argmax}_j \phi'_i x_{jt} + \psi_i c_{ijt} + \epsilon_{ijt},
 \end{aligned}$$

where U_{ijt} is the utility of alternative j for consumer i in period t ; V_{ijt} is the deterministic component of the utility of alternative j for consumer i in period t (i.e., $V_{ijt} = \phi'_i X_{jt} + \psi_i c_{ijt}$); X_{jt} is a vector of attributes for brand j in period t , (e.g., including price, brand dummies and other product characteristics); c_{ijt} is a latent indicator variable which is equal to 1 if consumer i has a coupon for brand j in period t , and 0, otherwise; ϕ_i and ψ_i are utility coefficients for consumer i , where the latter measures the utility trade-off for consumer i of using a coupon¹; and, ϵ_{ijt} is an individual-specific demand shock for the utility of alternative j for consumer i in period t .

Assuming ϵ_{ijt} is distributed according to the Extreme Value(0,1) distribution, the probability, p_{ijt} , that consumer i chooses brand j in period t is given by (Ben Akiva and Lerman,

¹One possible extension of the model is to specify different coupon coefficients for different brands (i.e., replace ψ_i by ψ_{ij}). In addition, if coupons with different face values were issued for a given brand, and if there is enough information for each of them, then one could also model the effect of different face values on the utility function of each consumer. Similarly, one could also model the effect of different distribution vehicles (e.g., on-pack versus in-pack) on consumer choice. These extensions can be easily incorporated to the methods that are introduced in this paper; however, they increase the computational requirements and, depending on how these extensions are modeled by the researcher, this may also reduce the degrees of freedom for estimating the response of consumers to coupon promotions.

1985):

$$(2) \quad p_{ijt}(c_{it}) \equiv P(y_{it} = j \mid c_{it}, \phi_i, \psi_i, x_t) = \frac{e^{\phi_i' x_{jt} + \psi_i c_{ijt}}}{\sum_{k=1}^J e^{\phi_i' x_{kt} + \psi_i c_{ikt}}},$$

which we specify explicitly as a function of c_{it} in order to emphasize the dependence of this probability (p_{ijt}) on the coupon indicator vector for consumer i in period t (c_{it}). Whenever this dependence is redundant, this choice probability will be simply denoted by p_{ijt} instead of $p_{ijt}(c_{it})$. For notational convenience, define z_{ijt} as a latent indicator variable equal to 1 if consumer i chooses brand j in period t (i.e., if $y_{it} = j$), and zero otherwise. In addition, let S_{jt} denote the observed aggregate market share of brand j in period t . Furthermore, assume that the researcher does not have access to individual-level data (i.e., z_{ijt} , c_{ijt}). Instead, the researcher only has aggregate information about coupon usage, coupon distribution and consumer choices (i.e., N_{jt}^c , N_{jt}^r and S_{jt}) from which inferences about consumer preferences ($\phi = \{\phi_i\}$, $\psi = \{\psi_i\}$) and coupon usage will be made.

According to these assumptions, the following restrictions must be satisfied by the unobserved (to the researcher) individual behavior of consumers in order to be exactly consistent with the observed aggregate data:

$$(3) \quad \sum_{i=1}^N z_{ijt} = NS_{jt} \quad (\text{Market Share}),$$

$$(4) \quad \sum_{i=1}^N c_{ijt} = N_{jt}^c \quad (\text{Coupon Distribution}),$$

$$(5) \quad \sum_{i=1}^N c_{ijt} z_{ijt} = N_{jt}^r \quad (\text{Coupon Redemption}).$$

Finally, the heterogeneity in consumer preferences is modeled by specifying that each vector of coefficients $\theta_i \equiv (\phi_i, \psi_i)'$ is independent and identically distributed according to a multivariate normal distribution with mean $\bar{\theta}$ and variance-covariance matrix D , as is common in empirical applications.

2.1 Likelihood Function and Posterior Density

Using a data augmentation strategy (Tanner and Wong 1987), the unobserved individual data about choices (z_{ijt}) and coupon usage (c_{ijt}) are treated as parameters (missing data, Little and Rubin 1987), which will be simulated from their posterior distribution. In order to formulate the posterior distribution, the likelihood of the augmented data is specified for this demand model:

$$(6) \quad \mathcal{L}_{aug} = \left(\prod_{i=1}^N \prod_{j=1}^J \prod_{t=1}^T p_{ijt}(c_{ijt})^{z_{ijt}} \right) \mathbb{I}_{\{(Z,C) \in \Omega\}},$$

where:

$$(7) \quad \Omega = \left\{ (Z, C) : \sum_{i=1}^N z_{ijt} = N S_{jt}, \sum_{i=1}^N c_{ijt} = N_{jt}^c, \sum_{i=1}^N c_{ijt} z_{ijt} = N_{jt}^r, \sum_{j=1}^J c_{ijt} \leq J \right\};$$

the indicator function ensures that the (augmented) individual choices and coupon variables are exactly consistent with the aggregate information according to equations (3), (4) and (5); and Ω is set of all possible configurations of choices and coupon usage (Z, C) satisfying these constraints. It is important to note that the use of indicator functions to incorporate restrictions on the parameters of a model has been previously proposed in the context of Bayesian estimation by Gelfand, Smith and Lee (1992) and it has been used, for example, by McCulloch and Rossi (1994) in their analysis of the multinomial probit model, where latent utilities are sampled from truncated normal distributions. Using equation (6), the posterior density of the parameters and the augmented data (Z, C) is proportional to the following expression:

$$(8) \quad f(Z, C, \theta, \bar{\theta}, D | S, N^c, N^r, X) \propto \left(\prod_{i=1}^N \phi(\theta_i; \bar{\theta}, D) \prod_{j=1}^J \prod_{t=1}^T p_{ijt}(c_{it})^{z_{ijt}} \right) \mathbb{I}_{\{(Z,C) \in \Omega\}} \pi(\bar{\theta}, D),$$

where θ is a matrix containing each of the vectors of individual coefficients (θ_i); S denotes

the observed data matrix with the market shares of each of the J alternatives in each period; X corresponds to a matrix containing marketing information for each of the J alternatives in each of T periods (e.g., prices and brand dummies); N^c and N^r are matrices specifying the total number of coupons and the number of redeemed coupons for each of the J alternatives in each of T periods; $\phi(\cdot; \bar{\theta}, D)$ is the density of a multivariate normal random variable with mean $\bar{\theta}$ and variance-covariance matrix D ; and $\pi(\bar{\theta}, D)$ is the hyperprior for $\bar{\theta}$ and D , which is specified here as a standard Normal-Inverse Wishart prior (see Gelman, Carlin, Stern and Rubin, 1995, p. 80).

After formulating the augmented likelihood and the posterior density, the next subsection describes the implementation of a Markov chain Monte Carlo (MCMC) method for the estimation of the parameters of the demand model. Specifically, for each of the parameters, random draws are generated from their full-conditional posterior distribution (Gibbs sampling) and these values are used to make posterior inferences about these parameters and, hence, assess the effectiveness of coupon promotions.

2.2 Estimation

As in Chen and Yang (2006) and Musalem, Bradlow and Raju (2006), the estimation method here relies on the fact that after conditioning on the current values of the individual choices and coupon usage variables, Z and C , the parameters $\{\theta_i\}_{i=1}^R$, $\bar{\theta}$ and D can be sampled using standard Bayesian methods (Allenby and Rossi 2003); for brevity not described here. Therefore, the discussion is focused on the problem of how to generate draws of the augmented individual choices (Z) and coupons (C).

The procedure proposed in this section generalizes the pair-switching Gibbs sampler introduced in Musalem, Bradlow and Raju (2006) which only considered one type of restriction (market share). Under the generalization presented here, the augmented individual choices and coupons for each period, (z_t, c_t) must satisfy the constraints defined in equations (3), (4) and (5). For computational convenience, (z_t, c_t) are simulated by first assigning consumers

to pairs and then sequentially updating the choices and coupons in each pair. We note that the number of all feasible instances of z and c for a pair of consumers can be as large as 2^J , which makes the computation of the full-conditional posterior probability more cumbersome. This complexity can be reduced by first drawing choices (Z) conditioning on imputed coupons and all other parameters, and then drawing coupons (C) conditioning on imputed choices and all other parameters. These two steps are described in the following subsections.

2.2.1 Drawing choices from their full-conditional posterior distribution

Suppose the choices of consumers i_1 and i_2 in period t are considered, conditioning on all other parameters including their coupon availability parameters and the choices and coupon availability of all other consumers. Then, using (8), the full-conditional posterior distribution of the choices of these two consumers in period t is given by:

$$(9) \quad f(z_{i_1 t}, z_{i_2 t} | *) = K \cdot \mathbb{I}_{\{(Z,C) \in \Omega\}} \prod_{j=1}^J p_{i_1 j t}(c_{i_1 t})^{z_{i_1 j t}} p_{i_2 j t}(c_{i_2 t})^{z_{i_2 j t}}$$

where K is a normalization constant that depends on the values of all other parameters and the observed aggregate data. Assuming that in a given iteration of the Gibbs sampler, the values of Z and C satisfy constraints (3), (4) and (5), it is easy to verify that when the choices of all other consumers are held constant, there are only two instances of $\{(z_{i_1 t}, z_{i_2 t})\}$ that may have a non-zero probability. The first corresponds to the current values of $\{(z_{i_1 t}, z_{i_2 t})\}$, while in the second instance consumers i_1 and i_2 interchange their choices in period t . Note that any other configuration would violate one or more of the constraints in (3). In addition, it is necessary to take into account that a change in the choices of these two consumers $(z_{i_1 t}, z_{i_2 t})$ may also affect the constraint related to the number of redeemed coupons in period t (because z is also present in equation (5)). Accordingly, the interchange of choices is only feasible when the total number of redeemed coupons for each brand is the same with and without interchanging $z_{i_1 t}$ by $z_{i_2 t}$.

Consequently, the full-conditional posterior probability of the event where the choices of these two consumers take their current values corresponds to:

$$(10) \quad f(z_{i_1t}, z_{i_2t} | *) = \frac{\prod_{j=1}^J p_{i_1jt}(c_{i_1t})^{z_{i_1jt}} p_{i_2jt}(c_{i_2t})^{z_{i_2jt}}}{\prod_{j=1}^J p_{i_1jt}(c_{i_1t})^{z_{i_1jt}} p_{i_2jt}(c_{i_2t})^{z_{i_2jt}} + \prod_{j=1}^J p_{i_1jt}(c_{i_1t})^{z_{i_2jt}} p_{i_2jt}(c_{i_2t})^{z_{i_1jt}}},$$

while the complement of this expression defines the probability of interchanging the choices of both consumers. Estimation details are presented in Appendix A.

2.2.2 Drawing coupons from their full-conditional posterior distribution

We specify a procedure for sampling the coupon variables (C) from their full-conditional posterior distribution that is conceptually very similar to the one proposed in §2.2.1. As before, suppose we consider the coupons of consumers i_1 and i_2 in period t focusing only on the coupon indicator variables for brand A (i.e., c_{i_1At} and c_{i_2At}). Assuming that the current values of C and Z satisfy constraints (3), (4) and (5), the only instances of c_{i_1At} and c_{i_2At} that satisfy the constraint on the total number of coupons (equation (4)) are the current values and the instance where these values are interchanged. In addition, it is also necessary to consider that a change in the coupons for brand A of these two consumers (c_{i_1At}, c_{i_2At}) may also affect the constraint related to the number of redeemed coupons in period t (equation (5)). Accordingly, the interchange of coupons of brand A is only feasible when the total number of redeemed coupons for this brand is the same with and without interchanging c_{i_1At} by c_{i_2At} . The details of this estimation method are also presented in Appendix A.

2.3 Identification

In this subsection, we provide an intuitive description of how changes in the observed aggregate data help identify the underlying latent individual-level patterns. A related and more detailed description involving an experimental design and simulation is given in Musalem,

Bradlow, and Raju (2006). As the identification of the distribution of consumer preferences has been previously studied by Bodapati and Gupta (2004) in a context in which the researcher has aggregate data about consumer choices (market share data), we focus our discussion on the additional complexity that arises from the use of aggregate data for one of the independent variables (coupon availability).

First, it is important to mention that from a Bayesian inference point of view and for any general estimation problem, as long as the researcher uses proper prior distributions on the hyperprior parameters, the corresponding posterior distribution of the parameters will be well-defined. In the case of the specific problem described in this section, it is evident that there are multiple sequences of individual choices and coupons that are exactly consistent with the aggregate data (i.e., in total there are $|\Omega|$ sequences that could have generated the aggregate data). Instead of trying to find the sequence of choices and coupons with the highest likelihood of having generated the aggregate data, the estimation approach relies on simulating a large number of these sequences, where each of them is sampled according to its posterior probability given the observed aggregate data. Therefore, the estimation approach is based on averaging our inferences about the demand parameters across multiple individual-behavior scenarios where each of them is weighted by its posterior probability.

An important issue to determine what information in the data identifies the parameters of the model is to understand how different distributions of coupon preferences generate different aggregate market share and coupon redemption data. In particular, a higher mean of the distribution of coupon preferences, holding everything else constant, induces a higher number of coupons redeemed and provides more incremental sales to the brands with a higher level of coupon availability. Moreover, a higher fraction of choices are associated with coupon redemption. In addition, an increase in the variance of the coupon coefficients has a smaller impact on the fraction of choices associated with coupon redemptions, given that a higher level of heterogeneity will lead to stronger coupon effects for some consumers, but weaker effects for others. However, this higher heterogeneity also implies more extreme preferences.

Therefore, consumers will react in a more consistent and stable way to coupon promotions than in a scenario with the same mean, but lower variance of consumer coefficients. Consequently, market shares in periods where coupons are available will exhibit lower variability. A simulation studying this issue is available upon request.

2.4 Numerical Example

In order to demonstrate the efficacy of this approach, a numerical example with $J = 3$ brands, $T = 50$ periods and $N = 500$ consumers is considered. The utility function of each of these consumers includes four explanatory variables. The first two correspond to brand dummies for the first two brands, the third is generated from a standard normal distribution and the fourth variable is a coupon indicator (c_{ijt}). The true mean and variance of the individual coefficients (θ_i) are chosen as $\bar{\theta} = (1 \ 1 \ -1 \ 1)'$ and $D = I_4$, respectively, where I_4 denotes the identity matrix with four rows and columns. In addition, coupons are randomly generated and the probability of a consumer receiving a coupon in a given period is equal to 0.1, 0.2 and 0.3 for brands 1, 2 and 3, respectively.

Using only aggregate information (i.e., market shares, number of redeemed coupons and number of coupons available for each brand in each period), $\bar{\theta}$ and D are estimated according to the procedure described in this section. The starting values for $\bar{\theta}$ and D correspond to $\bar{\theta} = (0 \ 0 \ 0 \ 0)'$ and $D = 0.1 \ I_4$. The starting values of the MCMC sampler for Z (choices) and C (coupons) are randomly chosen from a distribution that assigns the same probability to any configuration of choices and coupons satisfying constraints (3), (4) and (5). Specifically, this can be implemented by first assigning the choices of brand j in period t to NS_{jt} randomly chosen consumers, then, among those consumers, we randomly distribute N_{jt}^r redeemed coupons and finally, the remaining coupons N_{jt} are randomly distributed among the consumers not choosing brand j in period t . Finally, the following hyperprior distributions are specified: $\bar{\theta} \sim N(0, 10^5)$ and $D \sim \text{Inverse Wishart}(6, 6I_4)$, very weakly informative. The results are presented in Table 1 and they are based on a single run of

200,000 iterations, where only the last 100,000 were used for the estimation of $\bar{\theta}$ and D , the mean and variance of the preference coefficients².

== Insert Table 1 here ==

From the results in Table 1 it is observed that the true values of $\bar{\theta}$ and D are covered by their 95% posterior-probability intervals and that the estimated posterior means are very close to their true values, providing numerical support for the method introduced in this section.

3 Limited Information

In Section 2, it was assumed that the researcher had access to data on the number of redeemed coupons (N_{jt}^r) and the number of consumers that received a coupon (N_{jt}^c). The latter information on received coupons is invariably harder to obtain in practice. The extension presented here, which does *not* assume this knowledge, is therefore relevant for many practical applications. In contrast, one could try to estimate N_{jt}^c by using data on the number of distributed coupons. However, many factors contribute to make the effective number of consumers that have a coupon in a given period different from the original number of distributed coupons. For example, some coupons might never reach any consumers, some consumers may lose their coupons, or some consumers may realize that they could have used a coupon after its expiration date.

Accordingly, the estimation procedure is generalized by only requiring knowledge of the number of redeemed coupons, while the number of consumers that received a coupon will be estimated according to the methodology presented below³. Consequently, we replace **A2**

²Extensive simulation suggests that this was sufficient for convergence.

³As suggested by an anonymous reviewer, it might be argued that having access to coupon redemption data from multiple brands might still be difficult in practice. However, if a researcher obtains information from a retailer, then this assumption is likely to be reasonable, given that a Point-Of-Sale (POS) information system can easily record which transactions (or how many of them) were made using coupons for each brand in a product category. Nevertheless, an interesting avenue for future research is to generalize the

by:

(A2') Available Information: Redeemed Coupons. The researcher only observes aggregate data regarding the total number of redeemed coupons for each brand in each period while the total number of consumers that received a coupon is unknown to the researcher.

In addition, since it is possible that in some periods no coupons were available for a given brand (i.e., $c_{ijt} = 0$ for all i), we define a latent indicator variable δ_{jt} which is equal to 1 if coupons for brand j that are redeemable in period t were distributed, and it is equal to 0, otherwise. This new definition combined with **A2'** implies that we must replace constraint (4) by the following condition:

$$(11) \quad N_{jt}^r \leq \sum_{i=1}^N c_{ijt} \leq N\delta_{jt}.$$

We note that this last condition implies that when $N_{jt}^r > 0$, $\delta_{jt} = 1$. Therefore, only when $N_{jt}^r = 0$, there is uncertainty about δ_{jt} . In particular, if no coupons were redeemed, then there are two alternative explanations: i) no coupons were available ($\delta_{jt} = 0$), or, ii) coupons were distributed ($\delta_{jt} = 1$), but none of them was redeemed. In addition, we denote by r_{jt} the probability that a consumer will receive a coupon for brand j in period t , where r_{jt} is a function of δ_{jt} as shown below. We assume that each c_{ijt} is an independent Bernoulli random variable such that $P(c_{ijt} = 1) = r_{jt}$, for all i, j and t . In previous research (e.g., Erdem, Keane and Sun 1999), these coupon-availability probabilities (r_{jt}) have been estimated using disaggregate data assuming that they are constant across periods (i.e., $r_{jt} = r_j$). In this paper, we allow these probabilities to take different values in different periods and this is implemented by defining r_{jt} as follows:

$$(12) \quad r_{jt} = P(c_{ijt} = 1) = \delta_{jt} \frac{e^{\alpha_j + \nu_{jt}}}{1 + e^{\alpha_j + \nu_{jt}}}$$

current methodology to estimate the distribution of consumer preferences and coupon usage from aggregate information for only a subset of the brands.

where α_j is a fixed effect that determines the baseline probability of receiving a coupon for brand j in period t when $\delta_{jt} = 1$; and ν_{jt} is a zero-mean random effect that captures deviations from the baseline level (α_j). Furthermore, we specify the following prior and hyperprior distributions:

$$\begin{aligned}\delta_{jt} &\sim \text{Bernoulli}(q_j) \\ q_j &\sim \text{Beta}(a_j, b_j) \\ \alpha_j &\sim N(0, \sigma_j^2) \\ \nu_t &\sim \text{MVN}(0, \Sigma^c) \\ \Sigma^c &\sim \text{Inverse Wishart}(m_0, M_o).\end{aligned}$$

Finally, it is important to mention that the random effects for different brands ($\nu_{jt}, \nu_{j't}$) are allowed to be correlated via Σ^c . For example, a positive correlation would imply that when more coupons are available for one brand, more coupons are also available for the other brand (a consequence of many competitive coupon strategies).

3.1 Estimation

In this subsection, it is described how to simulate the unobserved coupon variables c_{ijt} from their full-conditional posterior distributions now given **A2'** rather than **A2**. Note that the simulation of the unobserved individual choices can be implemented using the same method described in the previous section, while all other parameters can be estimated using standard methods.

First note that in Section 2 it was not possible to update the coupons of a *single* consumer in a given period, holding the coupons and choices of all other consumers and all other parameters constant. The reason for this is that once the coupons of all other consumers are

held constant, there is only one value of the coupon variable for the corresponding consumer that satisfies condition (4). Moreover, if this was implemented, the coupon variables would remain at their initial values for every iteration of the Gibbs-sampler and, consequently, the Markov chain would not converge to the posterior distribution of the parameters.

In this section, however, since the equality constraint specified in (4) has been replaced by inequality (11) corresponding to **A2'**, it is possible to update the coupon variables of each consumer singly, conditioning on the coupons of all other consumers and all other parameters. In particular, a Metropolis-Hastings (MH) step is proposed where a candidate vector of coupons for a single consumer (c_{it}^*) is randomly generated from a distribution that assigns equal probability to every vector c_{it}^* satisfying constraints (5) and (11). The details of this procedure are presented in Appendix B.

3.2 Numerical Example

A numerical simulation example is constructed based on the same parameter values for $\bar{\theta}$ and D as in Subsection 2.4. The true values for the parameters that determine the coupon probabilities are specified as: $q = (0.4, 0.5, 0.6)$, $\alpha = (-2, -1, 0)$ and

$$\Sigma^c = \begin{bmatrix} 2.0 & 1.0 & -1.0 \\ 1.0 & 2.0 & 0.0 \\ -1.0 & 0.0 & 2.0 \end{bmatrix}.$$

A Beta(1,1) hyperprior distribution (i.e., Uniform(0, 1)) is used for q_1 , q_2 and q_3 (i.e., $a_j = b_j = 1$), a $N(0, 1000)$ for each α_j (i.e., $\sigma_j^2 = 1000$) and an Inverse Wishart(5, $5I_3$), weakly informative, for Σ^c . As mentioned before, only aggregate data on market shares and number of *redeemed* coupons for each brand in each period are used to estimate the posterior distribution of the parameters of the model (i.e., we do not use data on how many consumers received a coupon for each brand in each period and, of course, any individual-level data). The starting values for $\bar{\theta}$, D and Z are the same as those used Subsection 2.4, while the initial

values for α_j and Σ^c correspond to $\alpha_j = 0$ and $\Sigma^c = I_3$. In the case of the coupon variables (C), the initial value for N_{jt}^c is set to be equal to the integer part of $N_{jt}^r + 0.3(N - N_{jt}^r)$, and then these N_{jt}^c coupons are randomly assigned among the N consumers. Using the method proposed in Subsection 3.1 the results presented in Table 2 were obtained, where the results are again based on a single run of 200,000 iterations with the last 100,000 used for estimation.

== Insert Table 2 here ==

From the results it is observed that the true values of $\bar{\theta}$, D , α , q and Σ^c are covered by their 95% posterior-probability intervals and that the posterior means and posterior medians are very close to the true values (all within 1 posterior standard deviation).

4 Utility Maximization, Strategic Firm Behavior and Coupon Expiration

In this section, three extensions are introduced. These extensions are related to the assumptions about the decision to use a coupon, the strategic coordination of coupon promotions with other marketing variables (e.g., prices) and the expiration of coupons. In addition, we also extend the methodology to control for price endogeneity (see Technical Appendix 1 for the implementation details of each of these extensions)⁴.

4.1 Utility Maximization and Coupon Usage

In the preceding sections, the redemption process was characterized by using **A5** as a simplifying assumption. In this section **A5**, is replaced by a structural demand assumption (i.e., an assumption consistent with utility maximization). This assumption is stated as follows:

⁴It is important to note that our ability to take into account the endogeneity of coupon distribution and redemption depends on both the data available to the researcher and the appropriateness of the model and its parametric assumptions. For example, if the coupon distribution policy is based on a targeting approach, then the model should be modified to account for the endogeneity of coupon distribution and additional information may be necessary (e.g., demographic variables used under the targeting policy).

(A5') A consumer uses a coupon only if its redemption improves the utility of the chosen alternative.

Accordingly, we modify the formulation of the choice probabilities as follows:

$$(13) \quad p_{ijt}(c_{it}) = \frac{e^{\phi'_i x_{jt} + \max\{\psi_i, 0\} c_{ijt}}}{\sum_{k=1}^J e^{\phi'_i x_{kt} + \max\{\psi_i, 0\} c_{ikt}}}.$$

where, as before, ψ_i denotes the net utility of using a coupon which includes both the benefits and costs associated with a coupon redemption for consumer i . If $\psi_i < 0$ (for example, because of the associated hassle costs of clipping and redeeming a coupon), then even if the consumer has a coupon for the chosen brand, the consumer will decide not to use it. This approach is conceptually similar to the one used in Chiang (1995), where consumers are classified as coupon users or non-users. Furthermore, this new assumption also requires us to modify the formulation of the constraint associated with the total number of redeemed coupons. Consequently, equation (5) is replaced by the following expression:

$$(14) \quad \sum_{i=1}^N c_{ijt} z_{ijt} \mathbf{I}_{\{\psi_i > 0\}} = N_{jt}^r \quad (\text{Coupon Redemption})$$

where the indicator function $\mathbf{I}_{\{\psi_i > 0\}}$ is included in order to ensure that available coupons are redeemed only if they improve the utility of consumers. For notational convenience, define $c_{ijt}^r = c_{ijt} z_{ijt} \mathbf{I}_{\{\psi_i > 0\}}$.

In terms of estimation, the method to simulate coupons and choices from their posterior distribution requires only a few changes (see Technical Appendix 1). However, the updating of ψ_i (the net utility associated with the use of a coupon) is more complex, because now this parameter may satisfy one of the following constraints depending on a consumer's (augmented) behavior:

- a) Positive Truncation: If during the T time periods, a consumer decided to redeem a coupon at least once, then $\psi_i > 0$.

- b) Negative Truncation: If during any of the T time periods, the consumer had a coupon for her chosen brand, but she decided not to redeem the coupon, then $\psi_i < 0$.

The method for simulating ψ_i from its posterior distribution taking into account these constraints involves generating candidate draws from a truncated normal distribution (see Technical Appendix 1).

4.2 Strategic Firm Behavior: Coordination of coupon promotions with other marketing variables

The second generalization in this section is introduced in order to take into account the possibility that firms might strategically coordinate their coupon promotions with other marketing activities. For example, it has long been argued in the literature in marketing and economics that coupon promotions can be viewed as a price discrimination device. In the case of a monopoly that can target different segments of customers by setting different prices using coupons (third-degree price discrimination), regular prices are supposed to rise when coupons are offered in order to capture a higher revenue from non-users of coupons while still getting a fraction of the coupon users to buy the product at the discounted price. Consequently, the monopolist can collect higher profits by discriminating among coupon users and non-users. Anderson and Song (2004) have shown, however, that this is not necessarily true when coupon promotions are implemented as a form of second-degree price discrimination (i.e., coupons are available to all consumers, but consumers self-select and only those willing to use them will obtain the savings). In fact, in the case analyzed by Anderson and Song (2004), prices and coupons may exhibit a synergistic effect on profits and, under certain conditions, a firm might be better off by simultaneously lowering regular prices and offering coupons.

In order to capture this strategic coordination, we allow the coupon distribution policy to be a function of the pricing policy and other elements of the marketing mix. Specifically, it is

also assumed that the coupon distribution policy can be (locally) approximated by making the logit of the probability that a consumer receives a coupon for brand j in period t (i.e., $\ln(\frac{r_{jt}}{1-r_{jt}})$), a linear function of current prices and other marketing variables (as it will be shown below in equation (15)). Therefore, instead of imposing optimality conditions, the proposed method is based on directly estimating the coupon distribution policy, i.e., the extent to which coupon distribution and other marketing variables are coordinated. This policy estimation approach is related to the one used by Manchanda, Rossi and Chintagunta (2004) to infer, without imposing optimality conditions, the extent to which the level of sales force contact set by managers in the pharmaceutical industry is related to their knowledge about physician responsiveness⁵.

Accordingly, the expression for r_{jt} is formulated as follows:

$$(15) \quad r_{jt} = P(c_{ijt} = 1) = \delta_{jt} \frac{e^{\alpha_j + \rho' m_{jt} + \nu_{jt}}}{1 + e^{\alpha_j + \rho' m_{jt} + \nu_{jt}}}$$

where m_{jt} is a vector of marketing variables that are aligned with the coupon distribution policy for period t while ρ is the corresponding vector of coefficients. For example, the vector m_{jt} may include the prices of brand j . In particular, a positive effect of current prices on r_{jt} implies that an increase in the prices of a brand is likely to be aligned with a corresponding increase in the distribution of coupons for that brand.

Finally, in order to avoid estimation bias associated with price endogeneity (due to the possibility that prices might be set by firms with knowledge of market conditions unobserved to the researcher), a final extension is applied to the formulation of the utility function of consumers and its estimation. Specifically, *common* demand shocks are introduced in the utility function of consumers that capture temporal fluctuations in the demand of a product for reasons that are unobserved to the researcher (i.e., not associated with changes in the

⁵It is important to mention, however, that Manchanda, Rossi and Chintagunta (2004) deal with a very different endogeneity problem. In their case, they treat marketing activity (detailing) as a function of physician-specific responsiveness parameters. In this paper, however, a model of the relationship among different types of marketing variables is formulated (e.g., prices, feature and coupon promotions), although in both cases optimality conditions are not imposed on the firm behavior.

values of the marketing variables x_t):

$$(16) \quad p_{ijt}(c_{it}) = \frac{e^{\phi'_i x_{jt} + \max\{\psi_i, 0\} c_{ijt} + \xi_{jt}}}{\sum_{k=1}^J e^{\phi'_i x_{kt} + \max\{\psi_i, 0\} c_{ikt} + \xi_{kt}}}.$$

where ξ_{jt} denotes a common demand shock which modifies the utility of brand j for all consumers in period t . Common demand shocks play an important role in the model specification by preventing the model from becoming almost deterministic as the number of consumers generating the aggregate data increases (see Musalem, Bradlow and Raju (2006) for a discussion of this issue). Furthermore, given that market conditions in one period might be similar to those in previous weeks, we allow these common demand shocks to exhibit autocorrelation based on a first-order autoregressive model (AR(1)) such that:

$$(17) \quad \xi_{jt} = \gamma_{dj} \xi_{jt-1} + \tilde{\xi}_{jt}$$

where $\tilde{\xi}_{jt}$ is an independent demand shock for alternative j in period t and γ_{dj} is an autoregressive coefficient that captures the serial correlation in the common demand shocks for brand j . In order to control for the endogeneity of prices, each vector $\tilde{\xi}_t$ is assumed to follow a multivariate normal distribution with zero mean and is allowed to be correlated with prices. Specifically, following Yang, Chen and Allenby (2003), we model prices as a linear function of a vector of instruments (w_{jt}) and an error term (η_{jt}):

$$(18) \quad \text{price}_{jt} = w'_{jt} v_j + \eta_{jt},$$

where v_j is a vector of coefficients for brand j and the price shocks (η_{jt}) may also exhibit serial correlation according to an AR(1) model:

$$(19) \quad \eta_{jt} = \gamma_{pj} \eta_{jt-1} + \tilde{\eta}_{jt},$$

where $\tilde{\eta}_{jt}$ denotes the independent price shock for brand j in period t and γ_{pj} is an autoregressive coefficient that captures the serial correlation in the price shocks for brand j . Finally, each vector $\tilde{\eta}_t$ can be correlated with the corresponding vector of independent demand shocks $(\tilde{\xi}_t)$: $\begin{pmatrix} \tilde{\eta}_t \\ \tilde{\xi}_t \end{pmatrix} \sim N(0, \Sigma^d)$. As a result, this specification allows the vector of prices in a given period t (price_t) to be correlated with the contemporaneous vector of demand shocks (ξ_t) . Moreover, given that any AR(1) model can be represented as a moving average model of infinite order (MA(∞), Hayashi 2000, p.376), prices are also allowed to be correlated with non-contemporaneous demand shocks $(\xi_{t'}, t' \neq t)$, yielding a flexible specification to account for the potential endogeneity of prices.

4.3 Coupon Expiration

In the preceding sections, we assumed that coupons *immediately* expire after one period. If, however, the length of a period is equal, for example, to a week it might be reasonable to consider the possibility that if a coupon is not used in a certain period, the coupon might still be valid for redemption during the next period. Accordingly, we replace **A3** by **A3'**:

(A3') Non-Immediate Expiration. A coupon that has not been used in a given period, might still be valid for redemption during the next period.

We note that the possibility of redeeming a coupon in a future period obviously depends on factors that are unobserved to the researcher, such as whether the consumer will still hold that coupon in the next period and the expiration date of the coupon. In particular, the probability that a consumer has a valid coupon in period t , given that she had a coupon in $t-1$ that was not redeemed, is not necessarily equal to 1 (e.g., the coupon might have expired or the consumer might have lost the coupon). Consequently, we model this coupon carry-over effect by specifying different coupon probabilities depending on whether a consumer had a coupon in the previous period and whether that coupon was redeemed. These probabilities

for periods 2 to T are defined as follows:

$$(20) \quad r_{ijt} = P(c_{ijt} = 1) = \delta_{jt} \frac{e^{\alpha_j + \rho' m_{jt} + \alpha_{J+1} c_{ijt-1} (1 - c_{ijt-1}^r) + \nu_{jt}}}{1 + e^{\alpha_j + \rho' m_{jt} + \alpha_{J+1} c_{ijt-1} (1 - c_{ijt-1}^r) + \nu_{jt}}}, \quad 2 \leq t < T,$$

where c_{ijt-1}^r equals 1 if consumer i redeemed a coupon for brand j in period $t - 1$ (i.e., $c_{ijt-1}^r = c_{ijt-1} y_{ijt-1} I_{\{\psi_i > 0\}}$) and α_{J+1} allows the coupon-availability probability to change when a consumer had a coupon in the previous period which was not redeemed. For example, positive values of α_{J+1} imply that if a coupon was available to consumer i in period $t - 1$, but the consumer did not use it (i.e., if $c_{ijt-1} (1 - c_{ijt-1}^r) = 1$), then the probability that the consumer will have a valid coupon in period t is higher⁶.

In addition, it is also necessary to define coupon probabilities for the first period that do not depend on c_{ij0} or c_{ij0}^r , data usually unobservable. Hence, we instead specify a different model for r_{ij1} , which is defined as follows⁷:

$$(21) \quad r_{ij1} = P(c_{ij1} = 1) = \delta_{j1} \frac{e^{\alpha_j + \rho' m_{j1} + \varphi_j}}{1 + e^{\alpha_j + \rho' m_{j1} + \varphi_j}}.$$

where φ_j is a fixed effect associated with brand j in period 1. We note that φ_j will only be relevant if $\delta_{j1} = 1$. Therefore, only if the event that $\delta_{j1} = 1$ has significant (posterior) probability, will it be possible to estimate φ_j . If that probability is very small, then we can simply ignore φ_j for any practical purposes.

⁶As indicated by an anonymous reviewer, an alternative specification is to create a coupon stock variable based on the Nerlove-Arrow goodwill advertising model (Nerlove and Arrow 1962). In this context, the impact of coupons in the utility of a consumer for a given brand could be formulated as a function, for example, of all coupons non-previously redeemed. This alternative formulation, however, is more complex from a computational point of view, given that changes in coupon availability in one period, change the utility function of the consumer in all successive periods until the next redemption.

⁷We note that even for consumer panel data, we only observe C when there is a coupon redemption. When there is no redemption, we do not know whether the consumer had a coupon for a non-chosen alternative.

4.4 Numerical Example

A numerical simulation example is constructed where consumers choose among three brands and a no-purchase alternative. The utility function includes three brand intercepts and a covariate $(x_{4,jt})$ generated from a standard normal distribution. The true values for $\bar{\theta}$ and D correspond to $(1, 1, 1, -1, 1)$ and I_5 . In terms of the formulation of the coupon probabilities, m_{jt} includes $x_{4,jt}$ (the value of the fourth explanatory variable for brand j in period t) and we set $\alpha = (-2, -1, 0, 2)$, $\rho = 1$ and $\varphi = (1, 0, -1)$. In addition, we use the same true values for q and Σ^c as in Subsection 3.2. We constrain δ_{j1} to be equal to 1 for every brand in order to obtain meaningful estimation results for each of the components of γ (in our real data analyses in §5 this is not required). The coefficients capturing the autocorrelation in the demand and price shocks are defined as follows: $\gamma_p = (0.2, 0.0, -0.3)$ and $\gamma_d = (0.0, 0.3, -0.2)$. In addition, $v_1 = (0, 1)'$, $v_2 = (0, 2)'$, $v_3 = (0, 0.5)'$ and Σ^d is specified as follows:

$$\Sigma^d = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.30 & 0.20 & -0.20 \\ 0.00 & 1.00 & 0.00 & -0.20 & 0.30 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.20 & 0.30 \\ 0.30 & -0.20 & 0.00 & 0.25 & 0.00 & 0.00 \\ 0.20 & 0.30 & 0.20 & 0.00 & 0.25 & 0.00 \\ -0.20 & 0.00 & 0.30 & 0.00 & 0.00 & 0.25 \end{bmatrix}.$$

Finally, the hyperprior distributions for these parameters are $\alpha_{J+1} \sim N(0, 10)$, $\rho \sim N(0, 1000)$, $\varphi_j \sim N(0, 10)$, $v_j \sim N(0_2, 10^5 I_2)$, $\gamma_{pj} \sim N(0, 1)$, $\gamma_{dj} \sim N(0, 1)$ and $\Sigma^d \sim \text{Inverse Wishart}(5, 5 \cdot (0.1 I_6))$, while all other parameters have the same hyperprior distributions as in the numerical experiment in Section 3. Tables 3 and 4 present the results, which are based on a single run of 600,000 iterations with the last 200,000 used for estimation, a longer run than before as the model is more complicated. From the results it is observed that the true values of $\bar{\theta}$, D , q , α , φ , ρ , Σ^c , v , γ and Σ^d are covered by their 95% posterior-

probability intervals (except for v_{32} , Σ_{16}^d , Σ_{23}^d and Σ_{34}^d) and that the true values are in most cases within one posterior standard deviation around their posterior means. In addition, the posterior mean and median of all of the non-zero off-diagonal elements of Σ^d and the non-zero elements of γ have the right sign.

== Insert Tables 3 and 4 here ==

In summary, this simulation and those presented in the previous sections demonstrate the efficacy of the general methodology under the most basic to a more realistic set of conditions. This methodology is now applied to a real data set.

5 Empirical Application

In this section, the method described in Section 4 is applied to a data set of purchases in the ice cream product category⁸. In order to provide an empirical validation of this method, a data set for which disaggregate data are available is used and two separate estimation procedures are implemented: disaggregate and aggregate estimation. While normally this would not be available, applying both methods and comparing their results provides an extremely strong empirical test to validate the proposed methodology. Moreover, a series of posterior predictive checks (Gelman, Meng and Stern 1996) will be performed to compare the inferences obtained from the aggregate estimation about individual choices and coupon usage with their corresponding true values, which can be computed using the disaggregate data.

In the case of the disaggregate estimation, individual-level data on choices and coupon redemption for eight different ice cream brands (Baldwin, Breyers, Country Charm, Dean Foods, Dreyers Edys, Fieldcrest, Private Label, Sealtest) are used, which were generated by a panel of consumers at a single store in an urban market in the period between June 1992

⁸We thank IRI for making these data available.

and June 1994. A total of 165 panelists that made at least four purchases during the 99 weeks of observation were selected⁹. In the case of the aggregate estimation, only the total number of choices and coupons redeemed for each brand in each week is used.

In terms of the utility function, in both cases dummy variable for each brand (x_1, \dots, x_8), prices (x_9) and feature (x_{10}) are used as covariates. These last two variables (price and feature) are also included in m_t (see equation 15) to capture a potential coordination between these elements of the marketing mix and the coupon distribution efforts. In addition, a non-purchase option is included in the set of alternatives to estimate category expansion effects. Table 5 presents summary statistics for this data set.

== Insert Table 5 here ==

Finally, we used data on input prices (e.g., milk, sugar, corn syrup, cream, packaging, electricity) as instruments to control for price endogeneity (w_{jt} in equation 18). We note that the results obtained with input prices as instruments are very similar to the results obtained using lagged values of price and feature as instruments. Lagged values of price and feature are more strongly correlated with price, but could in principle be correlated with common demand shocks (e.g., if current demand is a function of marketing actions in previous periods). Consequently, the results presented here are based on the specification that uses input prices as instruments for price.

In order to assess the validity of these instruments, we estimated whether each of the instruments w_k is uncorrelated with the demand shocks of each brand (Duan and Mela 2006). This can be easily accomplished by using the MCMC draws of the product between each common demand shock ξ_{jt} and each instrument w_{kt} and estimating whether the mean of this product across time periods is significantly different from zero. Accordingly, none of these

⁹Given that the selected panelists made at least four purchases, the following constraint was added in the aggregate estimation: $\sum_{t=1}^T \sum_{j=1}^J z_{ijt} \geq 4$, for $i = 1, \dots, 165$. This constraint was included in order to make the results from both estimation procedures comparable. Note that this constraint can be easily incorporated in the Gibbs sampler by assigning zero probability to any draw that violates this inequality.

products was found to be significantly different from zero, and, consequently, we did not find evidence of a significant correlation between the instruments and the demand shocks.

Using the method presented in Section 4, we estimated the parameters of the demand and coupon availability model. We also estimated a constrained version of this formulation in order to assess the degree of generality needed in the model and to illustrate the model selection process. In particular, the constrained model considers a situation in which coupons expire immediately. For model comparison purposes, we computed the log-marginal likelihood of the data (presented in Table 6) given each of the two alternative models (full model and constrained model) under both estimation procedures (aggregate and disaggregate)¹⁰. From these results and according to the criterion in Kass and Raftery (1995) very strong empirical support is obtained for the second model (full model) under both estimation procedures (aggregate and disaggregate). In addition, we note that the fact that $\alpha_{Coupon_{t-1}}$ (the coefficient for the coupon carryover effect) is estimated to be significantly different from zero (see Table 10) also provides support for selecting the full model instead of the constrained model.

== Insert Table 6 here ==

The rest of our discussion is focused on the results obtained for the selected model (full model), which are reported in Tables 7, 8, 9 and 10. Results for the off-diagonal elements of D , Σ^d and Σ^c are reported as correlations (i.e., normalized by the product of the corresponding standard deviations) and they are only reported if these elements are estimated to be significantly different from zero for at least one of the estimation procedures (aggregate and disaggregate). According to these results, it is verified that the estimated 95% posterior probability intervals for all parameters overlap each other under both estimations (except for D_{66} , D_{88} , $D_{67}/\sqrt{D_{66}D_{77}}$ and $D_{69}/\sqrt{D_{66}D_{99}}$). In terms of the demand parameters (see Table

¹⁰In the case of the aggregate estimation of both models and denoting by A the aggregate data, $(\ln(p(A)) - \ln(|\Omega|))$ is reported instead of $\ln(p(A))$, because $|\Omega|$ is constant under both models, and therefore this term is irrelevant for model comparison purposes. See Technical Appendix 2 for details on the estimation of the marginal likelihood under the aggregate estimation procedures.

7), the estimated posterior means for $\bar{\theta}$ under both cases are fairly close (within 1 posterior standard deviation from each other). In the case of the variance of the preference coefficients (D), those corresponding to the brand intercepts of the last three brands and the price coefficient are estimated to be somewhat smaller under the aggregate estimation, while the opposite is observed for the corresponding variance of the coefficients of feature and coupon. Some deviations (although non significant except in two cases) are also observed for the off-diagonal elements of D , given that the posterior means of most of the off-diagonal elements of D are closer to zero under the aggregate estimation. This implies that the aggregate data provide less information about these off-diagonal elements than the disaggregate data. In addition, it is observed that the posterior standard deviations for $\bar{\theta}$ and D are higher, in general, in the case of the aggregate estimation, which reflects the fact that there is higher uncertainty about the demand parameters when the estimation is based only on aggregate data.

== Insert Tables 6, 7, 8, 9 and 10 here ==

From a managerial point of view, and as it has been suggested by previous research on aggregate estimation (e.g., Christen et al. 1997), a more relevant comparison of these results can be obtained by computing the sets of own- and cross-price elasticities under both estimation procedures (aggregate and disaggregate). Table 11 shows the estimated posterior mean (first block of results), 2.5%-ile (second block) and 97.5%-ile (third block) of the price elasticities under the aggregate and disaggregate estimation, respectively. These elasticities were computed assuming mean levels of prices, feature and coupon availability. These results show that the 95% posterior probability intervals for these two sets of elasticities overlap each other and their posterior means are very similar. Therefore, both of them would generate similar managerial recommendations for pricing purposes. For example, if a firm wants to evaluate the impact of raising prices by one percent (holding everything else constant), both estimation procedures (aggregate and disaggregate) would suggest similar outcomes.

== Insert Table 11 here ==

Also note that the estimation of demand parameters used to compute these elasticities takes into account the potential endogeneity of prices. Moreover, it is observed that the vectors of demand shocks and prices exhibit a significant correlation (see Table 9). For example, $\Sigma_{\xi_5\eta_5}^d$ is estimated to be negative which implies that prices are lower in periods of higher demand for brand 5. Note that this last result is consistent with economic theories about countercyclical pricing behavior such as collusion models (Rotemberg and Saloner 1986), procyclical demand elasticities (Warner and Barsky 1995) and loss-leader pricing (Chevalier, Kashyap and Rossi 2003).

In terms of the parameters related to the coupon probabilities (α , ρ , γ , Σ^c and q), a great degree of agreement is observed between the two sets of estimated values (see Table 10). In fact, the posterior means for each parameter under both estimations are within one posterior standard deviation from each other. Moreover, in terms of the coordination of coupon promotions with other marketing variables, ρ_{price} is estimated to be negative ($p < 0.07$), while ρ_{feat} is estimated to be positive ($p < 0.001$) under both aggregate and disaggregate estimation. These results imply that coupon promotions are aligned with lower prices and higher feature advertising. We note that this estimated relationship between prices and coupon distribution is not consistent with the price-discrimination argument for coupon promotions in Narasimhan (1984). Narasimhan (1984) argues that a firm can achieve higher profits by raising its regular price and distributing discount coupons that would be redeemed by the more price sensitive consumers. A necessary condition for this result, however, is to observe that consumers that are willing to redeem coupons are more price sensitive than the non-redeemers. However, in this empirical application, the estimated correlation between the price and coupon coefficients ($D_{price,coupon}$) was not significantly different from zero under both estimation settings (aggregate and disaggregate). Consequently, given this non-significant correlation, we do not expect to observe a positive association between coupon promotions and prices.

Instead, this empirical finding is consistent with Anderson and Song (2004) who argue that these two elements of the marketing mix may display a synergistic effect on profits¹¹. Accordingly, a firm may achieve higher profits by simultaneously distributing coupons and lowering its regular price. This finding is also consistent with the empirical results in Nevo and Wolfram (2002) for the breakfast cereal product category who argue instead that competitive reasons might explain this observed negative relationship between coupon promotions and pricing. In addition, these results also provide evidence of coupon promotion coordination across different brands. In particular, Σ_{34}^c is estimated to be significantly positive, which implies that when more coupons are available for brand 3, more coupons are also available for brand 4 (note that both brands are owned by the same manufacturer).

Finally, a series of posterior predictive checks of the results obtained from the aggregate estimation were performed, the strongest possible check of our approach. Specifically, the following statistics were computed from the imputed (Z, C) under the aggregate estimation and compared to the “truth” using the disaggregate data:

1. Total purchases: proportion of consumers making at least k purchases.
2. Penetration by brand: proportion of consumers choosing brand k at least once.
3. Number of different brands: proportion of consumers buying k different brands (during the 99 weeks of data).
4. Coupon redemption: proportion of consumers redeeming at least k coupons.
5. Coupon penetration: proportion of consumers redeeming a coupon for brand k at least once.

For each of these measures, the corresponding true values were computed using the disaggregate data and then compared with those estimated under the aggregate procedure.

¹¹We note that Anderson and Song (2004) also control for face value differences across coupon drops, while the models described in this paper do not include this variable.

The results are presented in Figure 1 where the solid line represents the true values and the other three lines represent the estimated 2.5% quantile, mean and 97.5% quantile of the posterior distribution of these quantities. From these posterior predictive checks it is observed that the true values of total purchases, penetration, coupon redemption and coupon penetration are, in general, within their 95% posterior-probability intervals. In the case of the third measure, number of different brands, the estimated values for 6 of the 8 levels are within their 95% posterior probability intervals. A brief discussion of some extensions to the demand model that could potentially improve the results is presented in Section 7.

In summary, it is concluded from these results that the aggregate procedure is doing a good job at estimating the unobserved individual data of coupons and choices, although it appears that there is still room for improvement.

== Insert Figure 1 here ==

6 Estimating the Value of Information and Evaluating Alternative Targeting Approaches

In this section, it is shown how the aggregate estimation results can be used to compute the economic value of purchasing information on individual choices and coupon usage for coupon targeting purposes. This is an important problem given that many times firms have the opportunity to collect or purchase individual-level data. Moreover, as in Besanko, Dubé and Gupta (2003), the aggregate estimation results can also be used to quantify the benefits of combining the aggregate data with information from a single purchase occasion through Bayesian updating. Finally, it is also possible to estimate the benefits that would be obtained from the implementation of simple targeting rules, such as giving coupons to all consumers that chose a particular brand in a given time period.

Specifically, we consider the following alternative systems of coupon distribution:

1. Random System: coupons are distributed among consumers at random and each consumer has the same probability of being among those who received a coupon.
2. Updating based on one observation: the preferences of a consumer are estimated combining the results from the aggregate estimation with information from a single purchase occasion, i.e., which alternative was chosen in a given week. We note that this may include observations corresponding to non-purchases and also that these two pieces of information (aggregate data and a single individual observation) are combined via Bayesian updating. Then we estimate the change in expected profits for each consumer associated with a coupon drop. Next, we rank consumers based on this measure in descending order. For example, if the coupon distribution is designed to reach 10% of the consumers, we consider the 10% of consumers with the highest (positive) change in expected profits. Similarly, at the 100% level of coupon distribution, all consumers are considered to estimate the corresponding incremental sales.
3. Updating based on one purchase occasion: the preferences of a consumer are estimated combining the results from the aggregate estimation with information about the last purchase of that consumer in the product category, i.e. the last ice cream brand purchased by a consumer. Then, as in the previous case, consumers are ranked in descending order based on their expected incremental profits.
4. Targeting System: the preferences of a consumer are estimated assuming the firm purchased $T=99$ weeks of individual data (full length of the time series). In order to estimate the performance of this system using only aggregate data, we use the simulated individual choices and coupon parameters to estimate the preferences of each of the simulated consumers. Then, as in the previous two cases, consumers are ranked in descending order based on their expected incremental profits.
5. Brand buyers: If it is assumed that a fixed number of consumers will receive coupons, those consumers that purchased brand 4 (without loss of generality) in the current week

will be the first to be considered under this coupon distribution rule. Given a certain level of coupon distribution (e.g., 50% of consumers), after the segment of buyers of that brand is covered, the remaining coupons (if any) are distributed at random among the non-buyers.

6. Category buyers: If any fixed number of consumers receiving coupons is assumed, those consumers that purchased any brand in the product category in the current week will be the first to be considered in this distribution. As in the previous case, after the segment of category buyers is covered, the remaining coupons (if any) are distributed at random among the non-buyers of ice cream.

In order to illustrate the performance of these systems, the gross profits of brand 4 (Dean Foods) were simulated under these alternative strategies assuming that coupons for other brands are not available, price and feature are at their mean levels and common demand shocks are equal to zero. In addition, it is assumed that if a single observation is available, this data point was observed during a week in which coupons for brand 4 were available to all consumers. In order to compute incremental profits, we assumed a coupon face value equal to 50 cents (the median value in our data) and a gross margin equal to 78.8 cents (which is consistent with the 25% gross margin reported by Dean Foods in a recent annual report). Finally, we note that a large variety of other policy simulations are also possible.

Accordingly, we computed the corresponding incremental profits for different levels of coupon distribution ranging from 0% of consumers receiving coupons to 100%¹². Figure 2 shows the results of this analysis, where the vertical axis shows the estimated incremental profits which were normalized by dividing them by the the expected profits assuming no coupons are distributed. From these results it is possible to visualize the potential benefits of using consumer preference information to assist the design of coupon promotions and distribution strategies. It is evident from this figure, that there is substantial variability in

¹²We note that this analysis does not take into account competitive effects. For a discussion of some of these issues see Shaffer and Zhang (1995); Corts (1998); Nevo and Wolfram (2002); Besanko, Dubé and Gupta (2003); and Anderson and Song (2004).

terms of the performance of each of these systems. First note that if coupons are distributed at random, then a 10% increase in the level of coupon distribution always exhibit the same expected benefits in terms of incremental profits. Therefore, if we combine two samples of 10% randomly chosen consumers, we obtain twice the benefits. Consequently, we observe a linear relationship between coupon distribution and incremental profits under the random system and we can treat this system as a baseline case. For example, if coupons are *randomly* distributed to 50% of consumers (Random System), the corresponding incremental profits are equal to half the value of the maximum incremental profits under this policy.

== Insert Figure 2 here ==

Next, we consider the targeting system which represents a situation in which the firm has purchased or collected individual data for targeting purposes. This coupon distribution policy clearly dominates the performance of all other systems. Moreover, it is observed that under this targeting system it is possible to achieve values reasonably close to the maximum incremental profits by distributing coupons to only 50% of consumers, a much more efficient outcome when compared with the benefits of distributing coupons at random. Note that this result is explained by the existence of a segment of customers that are very unlikely to change their preferences after receiving a coupon (approximately 50% of the consumers in the market) and, therefore, identifying the members of this segment provides an opportunity for a firm to implement more efficient coupon promotions.

Considering these two extremes that correspond to using disaggregate data versus not using any information, it is observed that the remaining four systems, which rely on a single data point or simple targeting rules, exhibit a performance that falls exactly between these two systems (random and targeting). For example, when the results of the aggregate estimation are combined with information about the last brand purchased by a consumer, then if 50% of coupons are distributed, the incremental profits correspond to 76% of the maximum incremental profits (a 52% gain compared to the performance of the random system). Furthermore, we also note that by using simple targeting rules such as giving

coupons to buyers of brand 4 or buyers of the product category, we can still achieve important efficiency gains (e.g., 28%-30% at the 50% level of coupon distribution) when compared to the baseline system. Moreover, in this particular example, these simple targeting rules (brand buyers and category buyers) exhibit a better performance than those that rely on combining aggregate data with a single data point for levels of coupon distribution below 30%.

In sum, the results presented in this subsection show that it is possible to make targeting methods, such as those proposed by Rossi, McCulloch and Allenby (1996), even more powerful given that one can estimate the distribution of consumer preferences from aggregate data and then use data, for example, corresponding to a single choice (e.g., the current transaction of the customer) to decide whether to give a coupon to a consumer using either Bayesian updating or even much simpler targeting rules.

7 Conclusions

In this paper, new methods have been presented for the estimation of demand models using only aggregate data on choices and coupon redemption. These methods allow researchers to specify models of consumer behavior and coupon usage at the individual-level and then estimate those models using only aggregate information. The main advantage of using models of individual behavior is that they can be easily derived and justified from theories of consumer behavior, such as random utility maximization. Consequently, the estimation results can be directly interpreted in terms of their implications for consumer behavior, as opposed to the results obtained from reduced-form models of aggregate behavior. These results can also be used to estimate the consequences of alternative coupon targeting strategies (policy experiments).

In terms of the estimation procedure, the method is based on simulating (data augmenting) the unobserved individual data taking into account: i) the probabilistic assumptions about the unobserved individual behavior of consumers, and ii) the aggregate information,

which is incorporated by specifying restrictions to the unobserved individual behavior. In this respect, the results from these methods may depend on the appropriateness of the assumptions specified by the researcher, as it is always the case for any empirical analysis. Consequently, future research should be aimed at generalizing the methods presented here in order to accommodate alternative specifications for the choice and coupon probabilities.

For example, potential problems associated with the IIA assumption at the individual level could be eliminated by replacing the multinomial logit model by a nested logit or a probit model. In addition, the demand model could be extended in order to consider effects on both primary and secondary demand (see Arora, Allenby and Ginter 1998). Other extensions include estimating different effects for each coupon distribution vehicle (e.g., in-pack, on-pack, peel-off; see Raju, Dhar and Morrison 1994) or for different face values (e.g., Krishna and Shoemaker 1992) and examining competitive effects (e.g., Besanko, Dubé and Gupta 2003). In any case, enabling the use of Bayesian methods when only aggregate data are available is a step forward for marketing researchers and their ability to make managerial decisions under a broader range of conditions.

Table 2: Results: Estimated posterior mean, standard deviation and quantiles for $\bar{\theta}$, D , q , α and Σ^c (limited information).

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	D_{11}	D_{22}	D_{33}	D_{44}	D_{12}	D_{13}	D_{14}	D_{23}	D_{24}	D_{34}
mean	1.064	1.008	-0.969	0.925	1.075	1.093	0.917	1.180	-0.011	0.057	0.109	-0.006	0.148	0.186
std.dev.	0.072	0.072	0.063	0.084	0.416	0.374	0.128	0.409	0.197	0.103	0.249	0.107	0.202	0.201
2.5%	0.933	0.876	-1.100	0.765	0.496	0.511	0.695	0.594	-0.373	-0.149	-0.388	-0.222	-0.259	-0.200
50.0%	1.061	1.005	-0.966	0.924	0.982	1.050	0.906	1.123	-0.024	0.058	0.117	-0.005	0.141	0.176
97.5%	1.218	1.161	-0.854	1.094	2.126	1.933	1.196	2.161	0.441	0.257	0.585	0.196	0.560	0.597
True Values	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000

	q_1	q_2	q_3	α_1	α_2	α_3	Σ_{11}^c	Σ_{22}^c	Σ_{33}^c	Σ_{12}^c	Σ_{13}^c	Σ_{23}^c
mean	0.408	0.560	0.655	-2.049	-1.095	0.182	1.704	2.397	1.584	0.8455	-0.278	-0.691
std.dev.	0.068	0.069	0.065	0.284	0.301	0.219	0.627	0.754	0.450	0.6192	0.380	0.524
2.5%	0.280	0.424	0.522	-2.597	-1.637	-0.264	0.864	1.336	0.939	-0.3089	-1.081	-1.776
50.0%	0.408	0.560	0.657	-2.058	-1.105	0.174	1.576	2.261	1.508	0.8163	-0.259	-0.681
97.5%	0.543	0.692	0.779	-1.477	-0.483	0.628	3.288	4.247	2.677	2.154	0.437	0.321
True Values	0.400	0.500	0.600	-2.000	-1.000	0.000	2.000	2.000	2.000	1.000	0.000	-1.000

Table 3: Results: Estimated posterior mean, standard deviation and quantiles for $\bar{\theta}$, D , α , ρ , φ , q and Σ^c (structural demand model).

	$\bar{\theta}_1$	$\bar{\theta}_2$	$\bar{\theta}_3$	$\bar{\theta}_4$	$\bar{\theta}_5$	D_{11}	D_{22}	D_{33}	D_{44}	D_{55}	D_{12}	D_{13}	D_{14}
mean	0.972	0.955	0.957	-1.086	1.067	1.641	1.287	1.968	1.183	1.137	-0.120	-0.091	-0.003
std.dev.	0.153	0.191	0.157	0.107	0.171	0.783	0.572	1.148	0.239	0.382	0.456	0.526	0.216
2.5%	0.664	0.538	0.620	-1.328	0.785	0.589	0.562	0.711	0.832	0.580	-0.931	-1.137	-0.449
50.0%	0.971	0.967	0.963	-1.076	1.049	1.464	1.162	1.675	1.145	1.066	-0.148	-0.098	0.013
97.5%	1.276	1.308	1.242	-0.906	1.456	3.512	2.826	5.179	1.774	2.009	0.919	1.102	0.402
True Values	1.000	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000

	D_{15}	D_{23}	D_{24}	D_{25}	D_{34}	D_{35}	D_{45}	α_1	α_2	α_3	α_4	ρ
mean	0.174	-0.085	0.009	-0.307	-0.308	0.071	0.037	-2.459	-1.014	-0.051	1.488	-0.837
std.dev.	0.349	0.412	0.215	0.331	0.252	0.401	0.223	0.594	0.379	0.358	0.976	0.113
2.5%	-0.490	-1.101	-0.430	-1.088	-0.907	-1.031	-0.437	-3.935	-1.766	-0.738	-0.363	-1.064
50.0%	0.146	-0.046	0.007	-0.298	-0.276	0.118	0.045	-2.399	-1.017	-0.060	1.472	-0.834
97.5%	0.909	0.669	0.416	0.305	0.078	0.714	0.485	-1.483	-0.259	0.683	3.548	-0.621
True Values	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-2.000	-1.000	0.000	2.000	-1.000

	φ_1	φ_2	φ_3	q_1	q_2	q_3	Σ_{11}^c	Σ_{22}^c	Σ_{33}^c	Σ_{12}^c	Σ_{13}^c	Σ_{23}^c
mean	1.311	0.444	-0.749	0.406	0.508	0.599	2.941	2.774	1.847	1.111	-0.500	-1.397
std.dev.	0.625	0.526	0.400	0.089	0.075	0.068	1.644	1.086	0.653	0.872	0.724	0.682
2.5%	0.238	-0.595	-1.539	0.252	0.362	0.464	1.221	1.321	0.955	-0.303	-1.831	-2.912
50.0%	1.258	0.448	-0.744	0.399	0.507	0.601	2.517	2.559	1.721	1.010	-0.539	-1.330
97.5%	2.792	1.480	0.035	0.603	0.654	0.727	7.340	5.479	3.483	3.165	1.136	-0.225
True Values	1.000	0.000	-1.000	0.400	0.500	0.600	2.000	2.000	2.000	1.000	0.000	-1.000

Table 4: Results: Estimated posterior mean, standard deviation and quantiles for v , Σ^d and γ (structural demand model).

	v_{11}	v_{12}	v_{21}	v_{22}	v_{31}	v_{32}	Σ_{11}^d	Σ_{22}^d	Σ_{33}^d	Σ_{44}^d	Σ_{55}^d	Σ_{66}^d	Σ_{12}^d	Σ_{13}^d
mean	0.610	1.044	0.618	2.093	0.377	0.306	1.461	1.442	0.725	0.379	0.399	0.353	-0.216	-0.213
std.dev.	0.208	0.111	0.166	0.131	0.095	0.087	0.315	0.308	0.159	0.117	0.127	0.114	0.219	0.161
2.5%	0.198	0.826	0.287	1.842	0.184	0.138	0.971	0.962	0.480	0.203	0.215	0.186	-0.673	-0.562
50.0%	0.611	1.044	0.617	2.091	0.379	0.305	1.419	1.402	0.703	0.361	0.377	0.334	-0.208	-0.204
97.5%	1.012	1.262	0.952	2.355	0.561	0.479	2.194	2.158	1.102	0.654	0.707	0.628	0.198	0.082
True Values	0.500	1.000	0.500	2.000	0.500	0.500	1.000	1.000	1.000	0.250	0.250	0.250	0.000	0.000

	Σ_{14}^d	Σ_{15}^d	Σ_{16}^d	Σ_{23}^d	Σ_{24}^d	Σ_{25}^d	Σ_{26}^d	Σ_{34}^d	Σ_{35}^d	Σ_{36}^d	Σ_{45}^d	Σ_{46}^d	Σ_{56}^d
mean	0.427	0.250	-0.456	0.340	-0.425	0.479	0.104	-0.206	0.176	0.231	-0.033	-0.123	-0.043
std.dev.	0.154	0.139	0.147	0.163	0.145	0.161	0.123	0.101	0.100	0.103	0.073	0.076	0.073
2.5%	0.175	0.011	-0.797	0.056	-0.755	0.220	-0.126	-0.429	0.003	0.055	-0.184	-0.299	-0.193
50.0%	0.410	0.239	-0.437	0.327	-0.407	0.459	0.099	-0.197	0.167	0.220	-0.033	-0.115	-0.041
97.5%	0.781	0.559	-0.221	0.704	-0.190	0.855	0.368	-0.029	0.397	0.462	0.112	0.006	0.102
True Values	0.300	0.200	-0.200	0.000	-0.200	0.300	0.000	0.000	0.200	0.300	0.000	0.000	0.000

	γ_{p1}	γ_{p2}	γ_{p3}	γ_{d1}	γ_{d2}	γ_{d3}
mean	0.175	-0.049	-0.360	-0.014	0.303	-0.176
std.dev.	0.112	0.116	0.171	0.142	0.121	0.152
2.5%	-0.052	-0.280	-0.708	-0.306	0.051	-0.483
50.0%	0.177	-0.048	-0.356	-0.011	0.307	-0.177
97.5%	0.392	0.173	-0.044	0.256	0.530	0.122
True Values	0.200	0.000	-0.300	0.000	0.300	-0.200

Table 5: Summary Statistics for the ice cream data.

Variable	Brand	Mean	Std. Dev.	Min.	Max.	Obs.
Market Share	1	0.006	0.011	0.000	0.055	$T=$ 99
	2	0.021	0.021	0.000	0.115	99
	3	0.011	0.016	0.000	0.085	99
	4	0.005	0.009	0.000	0.042	99
	5	0.008	0.020	0.000	0.133	99
	6	0.025	0.019	0.000	0.133	99
	7	0.016	0.020	0.000	0.103	99
	8	0.021	0.017	0.000	0.079	99
Prices	1	3.461	0.594	1.990	3.990	$T=$ 99
	2	3.162	0.410	1.690	3.490	99
	3	3.167	0.488	1.830	3.590	99
	4	3.152	0.489	1.740	3.590	99
	5	3.881	0.699	2.050	4.350	99
	6	1.609	0.141	0.850	1.690	99
	7	2.793	1.101	1.370	4.190	99
	8	2.580	0.180	1.990	2.890	99
Feature	1	0.171	0.378	0.000	1.000	$T=$ 99
	2	0.271	0.446	0.000	1.000	99
	3	0.183	0.378	0.000	1.000	99
	4	0.171	0.373	0.000	1.000	99
	5	0.126	0.312	0.000	1.000	99
	6	0.107	0.305	0.000	1.000	99
	7	0.117	0.283	0.000	1.000	99
	8	0.212	0.411	0.000	1.000	99
Coupon Redemption	1	0.162	0.889	0.000	8.000	$T=$ 99
	2	0.242	0.959	0.000	8.000	99
	3	0.505	1.955	0.000	14.000	99
	4	0.162	0.792	0.000	7.000	99
	5	0.222	1.374	0.000	13.000	99
	6	0.192	1.811	0.000	18.000	99
	7	0.475	1.913	0.000	13.000	99
	8	0.444	1.263	0.000	10.000	99

Table 6: Model Selection: Estimated Log-Marginal Likelihood

Model	Aggregate Estimation	Disaggregate Estimation
Constrained model	-14,023.5	-14,948.7
Full Model	-12,393.6	-13,354.7

Table 7: Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for $\bar{\theta}$ (mean of the demand coefficients) and the diagonal elements of D (variance of the demand coefficients; iterations 600,001-800,000).

Variable		Agg.				Disagg.			
		mean	(s.d.)	2.5%	97.5%	mean	(s.d.)	2.5%	97.5%
$\bar{\theta}$	Brand 1	-2.623	(0.727)	-4.302	-1.379	-2.524	(0.553)	-3.586	-1.391
	Brand 2	-0.568	(0.452)	-1.543	0.257	-1.013	(0.461)	-1.908	-0.048
	Brand 3	-1.925	(0.504)	-2.995	-0.970	-1.690	(0.452)	-2.582	-0.751
	Brand 4	-2.884	(0.689)	-4.399	-1.745	-2.725	(0.522)	-3.746	-1.687
	Brand 5	-1.892	(0.497)	-3.042	-0.945	-2.161	(0.589)	-3.592	-0.940
	Brand 6	-1.962	(0.273)	-2.516	-1.453	-2.546	(0.321)	-3.191	-1.907
	Brand 7	-1.914	(0.407)	-2.855	-1.231	-2.051	(0.348)	-2.734	-1.364
	Brand 8	-1.161	(0.398)	-1.896	-0.398	-1.466	(0.406)	-2.290	-0.645
	Price	-1.462	(0.120)	-1.700	-1.242	-1.452	(0.147)	-1.766	-1.160
	Feature	0.102	(0.162)	-0.237	0.400	0.225	(0.119)	-0.006	0.462
	Coupon	2.202	(0.561)	1.257	3.485	2.168	(0.368)	1.500	2.934
D	Brand 1	2.646	(1.284)	1.125	5.969	2.986	(0.963)	1.463	5.202
	Brand 2	1.920	(0.791)	0.845	4.035	2.727	(0.833)	1.390	4.616
	Brand 3	2.002	(0.784)	0.903	3.844	2.930	(0.885)	1.548	4.939
	Brand 4	2.443	(1.422)	0.900	6.607	2.915	(0.987)	1.372	5.179
	Brand 5	2.146	(0.857)	0.853	4.124	2.249	(0.744)	1.102	3.953
	Brand 6	1.420	(0.471)	0.731	2.559	5.644	(1.110)	3.795	8.151
	Brand 7	1.889	(0.884)	0.845	4.310	4.170	(0.970)	2.521	6.294
	Brand 8	1.598	(0.578)	0.783	3.013	5.370	(1.291)	3.189	8.288
	Price	0.278	(0.049)	0.199	0.389	0.502	(0.107)	0.330	0.745
	Feature	1.452	(0.473)	0.755	2.549	0.537	(0.098)	0.372	0.756
	Coupon	3.279	(1.412)	1.432	6.805	1.995	(0.559)	1.146	3.326

Table 8: Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles of the correlations of the demand coefficients ($D_{kl}/\sqrt{D_{kk}D_{ll}}$; iterations 600,001-800,000).

Variables	Agg.				Disagg.			
	mean	(s.d.)	2.5%	97.5%	mean	(s.d.)	2.5%	97.5%
Brands 1-6	0.048	(0.248)	-0.445	0.514	0.393	(0.169)	0.023	0.678
Brands 1-8	0.021	(0.261)	-0.464	0.515	0.486	(0.161)	0.106	0.739
Brand 1-Price	0.031	(0.184)	-0.329	0.386	-0.372	(0.160)	-0.639	-0.019
Brands 2-6	0.058	(0.250)	-0.429	0.534	0.363	(0.148)	0.046	0.621
Brand 2-Price	-0.181	(0.153)	-0.459	0.133	-0.392	(0.138)	-0.633	-0.103
Brands 3-4	0.030	(0.274)	-0.514	0.532	0.632	(0.126)	0.337	0.823
Brands 3-6	-0.027	(0.259)	-0.508	0.483	0.507	(0.132)	0.206	0.721
Brands 3-7	-0.044	(0.267)	-0.531	0.466	0.473	(0.150)	0.130	0.715
Brands 3-8	0.000	(0.258)	-0.484	0.489	0.554	(0.129)	0.264	0.761
Brand 3-Price	-0.038	(0.180)	-0.385	0.304	-0.505	(0.124)	-0.709	-0.226
Brands 4-6	0.063	(0.256)	-0.413	0.558	0.358	(0.162)	0.000	0.634
Brands 4-7	-0.059	(0.278)	-0.582	0.483	0.441	(0.153)	0.108	0.701
Brands 4-8	0.101	(0.234)	-0.345	0.559	0.536	(0.140)	0.223	0.759
Brand 4-Price	0.002	(0.188)	-0.361	0.358	-0.419	(0.142)	-0.660	-0.114
Brands 5-6	0.076	(0.232)	-0.381	0.542	0.391	(0.170)	0.021	0.679
Brands 5-7	0.104	(0.296)	-0.459	0.665	0.489	(0.162)	0.127	0.746
Brand 5-Price	-0.168	(0.169)	-0.475	0.166	-0.399	(0.157)	-0.657	-0.062
Brand 5-Coupon	0.045	(0.336)	-0.614	0.652	0.412	(0.160)	0.072	0.687
Brands 6-7	0.022	(0.254)	-0.468	0.507	0.743	(0.064)	0.600	0.849
Brands 6-8	-0.028	(0.241)	-0.468	0.443	0.510	(0.104)	0.284	0.688
Brand 6-Price	-0.216	(0.151)	-0.491	0.093	-0.659	(0.074)	-0.782	-0.494
Brands 7-8	-0.066	(0.230)	-0.492	0.386	0.414	(0.124)	0.148	0.630
Brand 7-Price	-0.203	(0.154)	-0.490	0.110	-0.596	(0.090)	-0.745	-0.400
Brand 8-Price	-0.119	(0.161)	-0.420	0.192	-0.582	(0.093)	-0.740	-0.378

Table 9: Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for the correlation of demand and price shocks $\Sigma_{kl}^d / \sqrt{\Sigma_{kk}^d \Sigma_{ll}^d}$ (iterations 600,001-800,000), only significant effects are presented.

Demand Shock	Price Shock	Agg.				Disagg.			
		mean	(s.d.)	2.5%	97.5%	mean	(s.d.)	2.5%	97.5%
Brand 1	Brand 3	0.457	(0.199)	0.038	0.798	0.458	(0.193)	0.051	0.788
Brand 1	Brand 4	0.469	(0.196)	0.054	0.802	0.474	(0.191)	0.072	0.796
Brand 5	Brand 5	-0.530	(0.202)	-0.850	-0.073	-0.692	(0.153)	-0.907	-0.325
Brand 6	Brand 7	0.362	(0.173)	0.002	0.676	0.382	(0.164)	0.040	0.681
Brand 8	Brand 2	0.366	(0.163)	0.032	0.666	0.345	(0.161)	0.013	0.641

Table 10: Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for the coupon availability parameters α , ρ , φ , Σ^c and q (iterations 600,001-800,000).

variable		Agg.				Disagg.			
		mean	(s.d.)	2.5%	97.5%	mean	(s.d.)	2.5%	97.5%
α	Brand 1	-6.334	(3.591)	-16.168	-0.389	-6.103	(2.835)	-11.956	-0.336
	Brand 2	-6.051	(1.782)	-9.550	-2.493	-5.848	(2.050)	-10.199	-2.005
	Brand 3	-5.539	(2.123)	-9.777	-1.524	-5.305	(1.989)	-9.321	-1.267
	Brand 4	-5.958	(2.569)	-11.187	-1.248	-5.343	(2.054)	-9.610	-0.995
	Brand 5	-5.075	(2.213)	-9.366	-0.875	-4.801	(2.019)	-8.331	-0.418
	Brand 6	-11.537	(3.104)	-17.746	-5.280	-9.453	(2.130)	-13.394	-4.871
	Brand 7	-5.342	(2.194)	-10.014	-1.254	-6.311	(2.473)	-11.287	-0.897
	Brand 8	-5.188	(1.828)	-8.812	-1.716	-4.939	(1.726)	-8.349	-1.200
	Coupon _{$t-1$}	8.513	(1.228)	6.024	10.702	7.670	(1.213)	5.253	9.772
ρ	Price	-1.082	(0.592)	-2.222	0.098	-0.941	(0.552)	-2.095	0.063
	Feature	3.684	(0.788)	2.229	5.257	3.180	(0.740)	1.835	4.888
φ	Brand 8	1.948	(0.861)	0.219	3.604	2.077	(0.923)	0.319	3.963
Σ_{jj}^c	Brand 1	11.913	(5.987)	4.507	27.768	9.058	(4.737)	3.049	20.885
	Brand 2	5.967	(2.582)	2.469	12.533	4.990	(2.188)	1.979	10.286
	Brand 3	11.258	(5.187)	4.641	24.998	8.519	(3.637)	3.751	17.714
	Brand 4	14.714	(8.349)	4.569	35.863	8.953	(4.428)	3.328	20.146
	Brand 5	6.127	(3.152)	2.024	14.264	4.211	(2.104)	1.657	9.605
	Brand 6	8.248	(4.481)	2.434	19.407	8.107	(5.172)	1.883	21.039
	Brand 7	12.641	(6.168)	4.069	27.368	10.428	(5.872)	3.233	26.426
	Brand 8	3.995	(1.628)	1.794	8.022	4.637	(2.082)	1.870	10.019
$\frac{\Sigma_{jk}^c}{\sqrt{\Sigma_{jj}^c \Sigma_{kk}^c}}$	Brands 1-3	0.579	(0.211)	0.085	0.889	0.544	(0.243)	-0.022	0.884
	Brands 1-4	0.578	(0.245)	0.023	0.926	0.478	(0.272)	-0.121	0.872
	Brands 3-4	0.639	(0.168)	0.249	0.882	0.627	(0.176)	0.191	0.866
q	Brand 1	0.685	(0.211)	0.263	0.988	0.723	(0.187)	0.318	0.988
	Brand 2	0.814	(0.151)	0.459	0.995	0.835	(0.128)	0.534	0.994
	Brand 3	0.805	(0.140)	0.488	0.993	0.840	(0.129)	0.526	0.994
	Brand 4	0.715	(0.177)	0.357	0.986	0.785	(0.169)	0.395	0.993
	Brand 5	0.738	(0.170)	0.374	0.986	0.696	(0.181)	0.338	0.983
	Brand 6	0.547	(0.268)	0.082	0.977	0.513	(0.263)	0.081	0.971
	Brand 7	0.476	(0.195)	0.179	0.923	0.629	(0.208)	0.236	0.977
	Brand 8	0.894	(0.088)	0.674	0.996	0.883	(0.092)	0.657	0.995

Table 11: Results: Estimated posterior mean, 2.5% and 97.5% quantiles of the own- and cross-price elasticities of the 8 brands in the ice cream product category (aggregate estimation).

		Agg.								Disagg.							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
mean	1	-3.40	0.07	0.03	0.02	0.01	0.04	0.03	0.06	-2.93	0.05	0.02	0.01	0.01	0.03	0.02	0.06
	2	0.02	-3.44	0.03	0.01	0.01	0.04	0.03	0.06	0.01	-2.61	0.02	0.01	0.01	0.03	0.01	0.04
	3	0.02	0.07	-3.22	0.01	0.02	0.04	0.03	0.06	0.01	0.06	-3.15	0.02	0.01	0.03	0.02	0.07
	4	0.02	0.07	0.03	-3.29	0.02	0.04	0.03	0.07	0.02	0.06	0.05	-3.17	0.01	0.03	0.03	0.07
	5	0.02	0.08	0.04	0.02	-3.94	0.04	0.04	0.07	0.01	0.05	0.02	0.01	-2.93	0.02	0.03	0.03
	6	0.01	0.06	0.02	0.01	0.01	-2.05	0.03	0.05	0.01	0.04	0.02	0.01	0.01	-2.57	0.05	0.06
	7	0.01	0.06	0.03	0.01	0.02	0.04	-3.12	0.05	0.01	0.04	0.02	0.01	0.01	0.09	-3.27	0.05
	8	0.02	0.07	0.03	0.01	0.01	0.04	0.03	-2.86	0.01	0.04	0.03	0.01	0.01	0.05	0.02	-3.34
2.5%	1	-4.44	0.02	0.01	0.00	0.00	0.02	0.01	0.02	-4.07	0.02	0.01	0.00	0.00	0.02	0.01	0.03
	2	0.01	-4.27	0.01	0.01	0.01	0.02	0.01	0.03	0.00	-3.52	0.01	0.00	0.00	0.02	0.01	0.02
	3	0.00	0.02	-4.15	0.00	0.01	0.02	0.01	0.02	0.00	0.02	-4.20	0.01	0.00	0.02	0.01	0.04
	4	0.00	0.03	0.01	-4.30	0.00	0.02	0.01	0.03	0.01	0.02	0.02	-4.17	0.00	0.01	0.01	0.04
	5	0.00	0.03	0.01	0.00	-5.02	0.02	0.01	0.03	0.00	0.02	0.01	0.00	-4.15	0.01	0.01	0.01
	6	0.00	0.03	0.01	0.00	0.00	-2.48	0.01	0.03	0.00	0.02	0.01	0.00	0.00	-3.10	0.03	0.04
	7	0.00	0.02	0.01	0.00	0.00	0.02	-3.88	0.02	0.01	0.02	0.01	0.01	0.00	0.05	-4.13	0.02
	8	0.01	0.03	0.01	0.01	0.01	0.02	0.01	-3.52	0.01	0.02	0.02	0.01	0.00	0.03	0.01	-4.21
97.5%	1	-2.21	0.16	0.09	0.05	0.04	0.09	0.06	0.13	-1.75	0.09	0.05	0.03	0.02	0.06	0.04	0.10
	2	0.04	-2.55	0.06	0.03	0.03	0.07	0.05	0.10	0.02	-1.55	0.04	0.02	0.02	0.05	0.02	0.06
	3	0.05	0.14	-2.03	0.04	0.04	0.07	0.06	0.12	0.03	0.10	-2.03	0.04	0.02	0.07	0.04	0.12
	4	0.06	0.15	0.08	-2.09	0.05	0.09	0.06	0.13	0.03	0.10	0.08	-2.06	0.02	0.05	0.05	0.13
	5	0.04	0.16	0.08	0.05	-2.65	0.09	0.09	0.14	0.03	0.10	0.05	0.02	-1.48	0.03	0.06	0.05
	6	0.03	0.10	0.04	0.03	0.02	-1.65	0.05	0.08	0.02	0.07	0.05	0.01	0.01	-2.10	0.08	0.10
	7	0.03	0.12	0.07	0.03	0.04	0.08	-2.06	0.10	0.03	0.06	0.05	0.03	0.03	0.14	-2.15	0.08
	8	0.03	0.11	0.06	0.03	0.03	0.06	0.05	-2.15	0.03	0.06	0.07	0.03	0.01	0.07	0.04	-2.55

Note: Cell entries (i, j) where i indexes row and j indexes column, give the percentage change in the market share of brand i corresponding to a 1% change in the price of brand j .

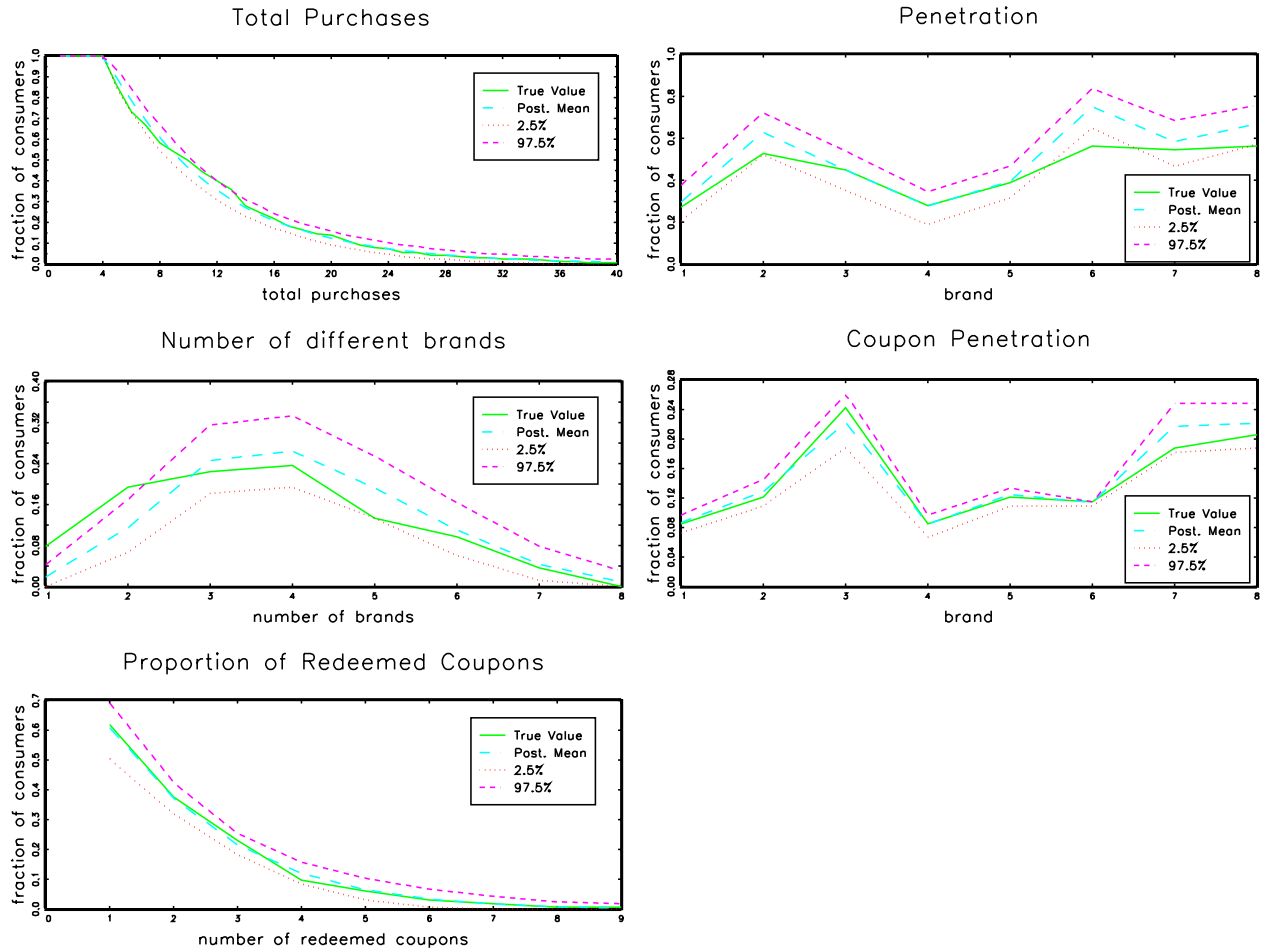


Figure 1: Posterior predictive checks: proportion of consumers making k or more purchases (total purchases); proportion of consumers that made at least one purchase for each brand (brand penetration); proportion of consumers buying k different brands (number of different brands); proportion of consumers redeeming at least k coupons (proportion of redeemed coupons); proportion of consumers redeeming a coupon for brand k at least once (coupon penetration).



Figure 2: Incremental profits for Brand 4 vs. Coupon Distribution (incremental profits are normalized by dividing them by the estimated profits under the scenario in which no coupons are distributed).

Appendix A: Sampling Choices and Coupons (basic model)

In this appendix we describe the procedure to sample choices and coupon availability parameters from their full-conditional posterior distribution according to the assumptions from Section 2.

A.1) Sampling Choices:

1. In each iteration (k) randomly select $N/2$ pairs of consumers without replacement and enumerate these pairs. Let (i_{1p}, i_{2p}) be the indexes of consumers in pair p and $(z_{i_{1p}t}^{(k)}, z_{i_{2p}t}^{(k)})$ their choices in period t in the current iteration k .
2. For each period t and starting from the first pair, successively and jointly draw the choices of each pair of consumers $(z_{i_{1p}t}^{(k+1)}, z_{i_{2p}t}^{(k+1)})$ from their full-conditional posterior distribution. Dropping the pair (p) and period (t) subscripts for notational convenience, we proceed by assigning $(z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, z_{i_1}^{(k)})$ according to the following probability:

$$(22) \quad f\left((z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, z_{i_1}^{(k)}) \mid *\right) = \frac{\prod_{j=1}^J p_{i_{1j}}^{z_{i_{2j}}^{(k)}} p_{i_{2j}}^{z_{i_{1j}}^{(k)}} \mathbb{I}_{\{c_{i_{1j}} z_{i_{1j}}^{(k)} + c_{i_{2j}} z_{i_{2j}}^{(k)} = c_{i_{1j}} z_{i_{2j}}^{(k)} + c_{i_{2j}} z_{i_{1j}}^{(k)}\}}}{\prod_{j=1}^J p_{i_{1j}}^{z_{i_{2j}}^{(k)}} p_{i_{2j}}^{z_{i_{1j}}^{(k)}} + \prod_{j=1}^J p_{i_{1j}}^{z_{i_{1j}}^{(k)}} p_{i_{2j}}^{z_{i_{2j}}^{(k)}}},$$

otherwise, let these choices remain at their current values by assigning: $(z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_1}^{(k)}, z_{i_2}^{(k)})$. Note that the indicator function keeps the total number of redeemed coupons constant.

Finally, the full-conditional posterior probability in equation (22) can be rewritten as follows:

$$(23) \quad f\left((z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, z_{i_1}^{(k)}) \mid *\right) = \frac{\mathbb{I}_{\{c_{i_{1j}} z_{i_{1j}}^{(k)} + c_{i_{2j}} z_{i_{2j}}^{(k)} = c_{i_{1j}} z_{i_{2j}}^{(k)} + c_{i_{2j}} z_{i_{1j}}^{(k)}\}}}{1 + \frac{\prod_{j=1}^J p_{i_{1j}}^{z_{i_{2j}}^{(k)}} p_{i_{2j}}^{z_{i_{1j}}^{(k)}}}{\prod_{j=1}^J p_{i_{1j}}^{z_{i_{1j}}^{(k)}} p_{i_{2j}}^{z_{i_{2j}}^{(k)}}}}.$$

A.2) Sampling Coupons:

1. In each iteration (k) randomly select $N/2$ pairs of consumers without replacement and enumerate these pairs. Let (i_{1p}, i_{2p}) be the indexes of consumers in pair p and $(c_{i_{1p}t}^{(k)}, c_{i_{2p}t}^{(k)})$ their coupon indicator vectors in period t in the current iteration k .
2. For each period t , each brand b and starting from the first pair, successively and jointly draw the coupons of each pair of consumers $(c_{i_{1p}bt}^{(k+1)}, c_{i_{2p}bt}^{(k+1)})$ from their full-conditional posterior distributions. Dropping the pair (p) and period (t) subscripts, this is implemented as follows:

- (a) Denote by $c_{i_1}^*$ a vector of coupon indicator variables such that $c_{i_1b}^* = c_{i_2b}^{(k)}$ and $c_{i_1b'}^* = c_{i_1b'}^{(k)}$ for all $b' \neq b$. Similarly, define $c_{i_2}^*$ a vector of coupon indicator variables such that $c_{i_2b}^* = c_{i_1b}^{(k)}$ and $c_{i_2b'}^* = c_{i_2b'}^{(k)}$ for all $b' \neq b$.

(b) Assign $(c_{i_1b}^{(k+1)}, c_{i_2b}^{(k+1)}) = (c_{i_2b}^*, c_{i_1b}^*)$ according to the following probability:

$$\begin{aligned}
f\left((c_{i_1b}^{(k+1)}, c_{i_2b}^{(k+1)}) = (c_{i_2b}^*, c_{i_1b}^*) \mid *\right) \\
&= \frac{\prod_{j=1}^J p_{i_1j}(c_{i_1}^*)^{z_{i_1j}} p_{i_2j}(c_{i_2}^*)^{z_{i_2j}} I_{\{c_{i_1j}^{(k)} z_{i_1j} + c_{i_2j}^{(k)} z_{i_2j} = c_{i_1j}^* z_{i_1j} + c_{i_2j}^* z_{i_2j}\}}}{\prod_{j=1}^J p_{i_1j}(c_{i_1}^*)^{z_{i_1j}} p_{i_2j}(c_{i_2}^*)^{z_{i_2j}} + \prod_{j=1}^J p_{i_1jt}(c_{i_1t}^{(k)})^{z_{i_1j}} p_{i_2j}(c_{i_2}^{(k)})^{z_{i_2j}}}
\end{aligned}
\tag{24}$$

otherwise, let these coupons for brand b remain at their current values by assigning: $(c_{i_1b}^{(k+1)}, c_{i_2b}^{(k+1)}) = (c_{i_1b}^{(k)}, c_{i_2b}^{(k)})$.

Appendix B: Sampling Coupons (limited information)

In this Appendix we describe the procedure to sample coupons from their full-conditional posterior distribution according to the assumptions from Section 3.

1. In every iteration k , for each period t and for every consumer i , successively draw $c_{it}^{(k+1)}$ as follows:
 - (a) Let b_i the brand chosen by consumer i in period t (i.e., $z_{ib_it} = 1$).
 - (b) Let c_{it}^* be such that:
 - i. $c_{ib_it}^* = c_{ib_it}^{(k)}$ (this condition is required in order to satisfy condition (5)), and,
 - ii. If $\delta_{b't} = 1$, generate $c_{ib't}^*$ from a Bernoulli distribution with probability 0.5, for all $b' \neq b_i$; otherwise, set $c_{ib't}^* = 0$, where the value of 0.5 was chosen in order to construct a symmetric jumping kernel¹³.
 - (c) Accept c_{it}^* , according to the following MH probability that takes into account the likelihood of coupons and choices:

$$\tag{25} \quad P\left(c_{it}^{(k+1)} = c_{it}^*\right) = \frac{\prod_{j=1}^J p_{ijt}(c_{it}^*)^{z_{ijt}} r_{jt}^{c_{ijt}^*} (1 - r_{jt})^{1 - c_{ijt}^*}}{\prod_{j=1}^J p_{ijt}(c_{it}^{(k)})^{z_{ijt}} r_{jt}^{c_{ijt}^{(k)}} (1 - r_{jt})^{1 - c_{ijt}^{(k)}}},$$

otherwise, assign $c_{it}^{(k+1)} = c_{it}^{(k)}$.

¹³One might be able to find other values for this probability that may induce a more efficient sampling of coupons from the posterior distribution. For example, one could potentially use the value of r_{jt} in the current iteration to generate a candidate vector of coupon indicator variables (c_{it}^*) .

References

- [1] ALLENBY, GREG M. AND PETER E. ROSSI (2003), “Bayesian Statistics and Marketing,” *Marketing Science*, 22 (3): 304-328.
- [2] ANDERSON, ERIC T. AND INSEONG SONG (2004), “Coordinating Price Reductions and Coupon Events,” *Journal of Marketing Research*, 41 (November): 411-422.
- [3] ARORA, NEERAJ, GREG M. ALLENBY AND JAMES L. GINTER (1998), “A Hierarchical Bayes Model of Primary and Secondary Demand,” *Marketing Science*, 17 (1): 29-44.
- [4] BAWA, KAPIL AND ROBERT W. SHOEMAKER (1987), “The Coupon-Prone Consumer: Some Findings Based on Purchase Behavior Across Product Classes,” *Journal of Marketing*, 51 (October): 99-100.
- [5] —, SRINI S. SRINIVASAN AND RAJENDRA K. SRIVASTAVA (1997), “Coupon Attractiveness and Coupon Proneness: A Framework for Modeling Coupon Redemption,” *Journal of Marketing Research*, 34 (November) : 517-525.
- [6] BEN AKIVA, MOSHE AND STEVEN R. LERMAN (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- [7] BERRY S. 1994. Estimating Discrete-Choice models of Product Differentiation. *RAND Journal of Economics* 25 (2): 242-262.
- [8] BERRY, STEVEN, JAMES LEVINSOHN AND ARIEL PAKES (1995), “Automobile Prices in Market Equilibrium,” *Econometrica*, 63 (4): 841-890.
- [9] BESANKO, DAVID, JEAN-PIERRE DUBÉ AND SACHIN GUPTA (2003), “Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Data,” *Management Science*, 49 (9): 1121-1138.
- [10] CHEN, YUXIN AND SHA YANG (2006), “Estimating Disaggregate Model Using Aggregate Data via Augmentation of Individual Choice,” *Journal of Marketing Research*, forthcoming.

- [11] CHEVALIER, JUDITH A., ANIL K. KASHYAP AND PETER E. ROSSI (2003), “Why Dont Prices Rise During Periods of Peak Demand? Evidence from Scanner Data,” *American Economic Review*, 93 (1): 15-37.
- [12] CHIANG, JEONGWEN (1995), “Competing Coupon Promotions and Category Sales,” *Marketing Science*, 10 (Fall): 297-315.
- [13] CHRISTEN, MARCUS, SACHING GUPTA, JOHN C. PORTER, RICHARD STAELIN AND DICK R. WITTINK (1997), “Using Market-Level Data to Understand Promotion Effects in a Non-Linear Model,” *Journal of Marketing Research*, 34 (3): 322-334.
- [14] DUAN, J. AND C. MELA (2006), “The Role of Spatial Demand on Outlet Location and Pricing,” working paper, Fuqua School of Business, Duke University.
- [15] ERDEM, TULIN, MICHAEL P. KEANE AND BAOHONG SUN (1999), “Missing price and coupon availability data in scanner panels: correcting for the self-selection bias in choice model parameters,” *Journal of Econometrics*, 89: 177-196.
- [16] GELFAND, ALAN E., ADRIAN F.M. SMITH AND TAI-MING LEE (1992), “Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling,” *Journal of the American Statistical Association*, 87 (418): 523-532.
- [17] GELMAN, ANDREW, JOHN B. CARLIN, HAL S. STERN AND DONALD B. RUBIN (1995), *Bayesian Data Analysis*. London: Chapman & Hall/CRC.
- [18] —, XIAO-LING MENG AND HAL S. STERN (1996), “Posterior predictive assessment of model fitness via realized discrepancies,” *Statistica Sinica*, 6, 733–807.
- [19] HAYASHI (2000), *Econometrics*. Princeton, NJ: Princeton University Press.
- [20] KASS, ROBERT E. AND ADRIAN E. RAFTERY (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 90 (430) : 773-795.
- [21] KRISHNA ARADHNA K. AND ROBERT W. SHOEMAKER (1992), “Estimating the effects of Higher Coupon Face Values on the Timing of Redemptions, the Mix of Coupon Redeemers and Purchase Quantity,” *Psychology and Marketing*, 9 (6) : 453-467.

- [22] LENK, PETER (1992), "Hierarchical Bayes Forecasts of Multinomial Dirichlet Data Applied to Coupon Redemptions," *Journal of Forecasting*, 11 : 603-619.
- [23] LEONE, ROBERT P. AND SRINI S. SRINIVASAN (1996), "Coupon Face Value: Its impact on coupon redemptions, brand sales and brand profitability," *Journal of Retailing*, 72 (3): 273-289.
- [24] LITTLE, RODERICK J.A. AND DONALD B. RUBIN (1987), *Statistical Analysis with Missing Data*, New York: John Wiley.
- [25] MANCHANDA, PUNEET, PETER E. ROSSI AND PRADEEP K. CHINTAGUNTA (2004), "Response Modelling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41 (November) : 467-478.
- [26] MCCULLOCH, ROBERT E. AND PETER E. ROSSI (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64: 207-240.
- [27] MUSALEM ANDRÉS, ERIC T. BRADLOW AND JAGMOHAN S. RAJU (2006), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," Working Paper, University of Pennsylvania.
- [28] NARASIMHAN, CHAKRAVARTHI (1984), "A Price Discrimination Theory of Coupons," *Marketing Science*, 3 (2): 128-147.
- [29] NERLOVE, MARC AND KENNETH J. ARROW (1962) "Optimal Advertising Policy under Dynamic Conditions," *Economica*, 39 (114): pp. 129-142.
- [30] NESLIN, SCOTT A. (1990), "A market response model for coupon promotions," *Marketing Science*, 9 (Spring): 125-145.
- [31] NEVO, AVIV AND CATHERINE WOLFRAM (2002), "Why do manufacturers issue coupons? An empirical analysis of breakfast cereals," *Rand Journal of Economics*, 33 (2): 319-339.
- [32] RAJU, JAGMOHAN S., SANJAY K. DHAR AND DONALD G. MORRISON (1994), "The Effect of Package Coupons on Brand Choice," *Marketing Science*, 13 (2): 145-164.

- [33] REIBSTEIN, DAVID J. AND PHILLIS A. TRAVER (1982), “Factors Affecting Coupon Redemption Rates,” *Journal of Marketing*, 46 (Fall): 102-113.
- [34] ROMEO, CHARLES J. (2007) “A Gibbs sampler for mixed logit analysis of differentiated product markets using aggregate data,” *Computational Economics*, 29 (1): 33-68.
- [35] ROTEMBERG, JULIO AND GARTH SALONER (1986), “A Supergame-Theoretic Model of Price Wars during Booms,” *American Economic Review*, 76 (3): 390-407.
- [36] ROSSI, PETER E., ROBERT E. MCCULLOCH AND GREG M. ALLENBY (1996), “The Value of Purchase History Data in Target Marketing,” *Marketing Science*, 15: 321-340.
- [37] TANNER, MARTIN A. AND WING H. WONG (1987), “The calculation of Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, 82: 528-550.
- [38] WARNER, ELIZABETH J. AND ROBERT B. BARSKY (1995), “The Timing and Magnitude and Retail Store Markdowns: Evidence from Weekends and Holidays,” *Quarterly Journal of Economics* 110 (2): 321-352.
- [39] YANG, SHA, YUXIN CHEN AND GREG M. ALLENBY (2003), “Bayesian Analysis of Simultaneous Demand and Supply,” *Quantitative Marketing and Economics* 1 (3): 251-275.