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

Andres Musalem Eric T. Bradlow Jagmohan S. Raju*

April 2005

*Andres Musalem is a doctoral candidate, The Wharton School of the The University of Pennsylvania and Instructor, Industrial Engineering Department of the University of Chile. Eric T. Bradlow is Associate Professor of Marketing and Statistics and Academic Director of the Wharton Small Business Development Center and Jagmohan S. Raju is Joseph J. Aresty Professor of Marketing, The Wharton School of the University of Pennsylvania. The authors thank David Bell for generously providing access to the data used in the empirical application of this paper. Please address all correspondence on this manuscript to Andres Musalem, 700 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340, Phone: (215) 746-0271, Fax: (215) 898-8243, amusalem@wharton.upenn.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 the study of 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, which differ in terms of their assumptions about consumer behavior and information availability, are analyzed.

Initially, we consider a situation where the researcher observes aggregate market shares, marketing activity, number of distributed coupons redeemed and the number of coupon redemptions for each brand in each period. In addition, we assume that each consumer may have a coupon for at most one brand in each period and coupons are only valid for one period. Then, we generalize the estimation procedure to handle more realistic situations. These generalizations include: i) each customer may simultaneously hold coupons for more than one brand, ii) the researcher observes the number of redeemed coupons in each period, but not the total number of consumers that received a coupon, and iii) a coupon that is not redeemed in one period can potentially be redeemed in future periods.

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.



*Key Words:* *Coupon Promotions, Discrete Choice Models, Random Coefficients, Data Augmentation, Markov Chain Monte Carlo.*

# 1 Introduction

Consumer packaged goods (CPG) manufacturers invest billions of dollars every year in coupon promotions. According to the Promotion Marketing Association, CPG manufacturers spent $7 billion in coupons and issued 258 billion coupons in 2003, from which only 3.6 billion coupons were redeemed (i.e, an average redemption rate of 1.5%). 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 who redeems coupons and who does 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).

In addition, the use of disaggregate data has also been very common among researchers using secondary data (i.e., consumer panel data) to estimate the behavioral and managerial implications of coupon promotions (e.g., Ward and Davis 1978; Narasimhan 1984; Bawa and Shoemaker 1987; Neslin 1990; Chiang 1995; Leone and Srinivasan 1996; Erdem, Keane and Sun 1999; Bell and Chiang 2001), or to make these inferences from the estimation of 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), while other researchers have only focused on coupon redemption without explicitly modelling brand choice (e.g., Reibstein and Traver 1982; Lenk 1992), have used reduced-form models to study the efficiency of coupon promotions (e.g., Anderson and Song 2004), or have estimated the impact of coupons by treating them as price reductions (e.g., Besanko, Dubé and Gupta 2003).

Recent advances in the Bayesian analysis of aggregate data initially proposed by Chen and Yang (2004) and then extended by Musalem, Bradlow and Raju (2004), have provided 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, we present in this paper a new methodology that is based on augmenting the observed aggregate data (market shares, number of redeemed coupons) with unobserved (simulated) sequences of choices *and* coupon usage. This methodology allows us to estimate the distribution of preferences among consumers and, consequently, the impact of coupons on the sales of each brand using only aggregate data with methods normally reserved for consumer panel coupon data sets.

Several scenarios are analyzed, which differ in terms of the assumptions about consumer behavior and information availability. In the simplest case, we consider a situation where the researcher observes the market share, marketing activity, number of coupons redeemed and number of consumers holding a coupon for each brand in each period. In addition, we assume that each consumer may have a coupon for at most one brand in each period and that coupons are only valid for one period. This might be reasonable when manufacturers do not promote their brands very often using coupons and when the duration of these coupon promotions is relatively short. We note, however, that in many practical contexts these conditions may not be appropriate and, consequently, we generalize this estimation procedure in order to handle more realistic situations. These generalizations include:

1. Multiple coupons: a customer may simultaneously hold coupons for more than one brand.

2. 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.

3. Non-expiring coupons: a coupon that is not redeemed in one period can potentially be

redeemed (with some probability) in the next period.

The proposed methods allow us 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 M 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 we can not only compute a point estimate, but also estimate its entire posterior distribution using Markov Chain Monte Carlo (MCMC) simulation, our 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, our methodology allows us to simulate the effects of policy experiments, such as reducing the number of coupons that are distributed, or reallocating the number of coupons distributed by a manufacturer across several brands.

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. Using this methodology we are able to account for heterogeneity in consumer preferences and it is possible to answer relevant managerial questions and analyze the consequences of different couponing strategies.

The rest of this paper is organized as follows. In Section 2 we describe and analyze a simple case that illustrates the basic ideas of the proposed methodology and introduces the needed notation. In Section 3 we generalize the estimation procedure by allowing consumers to simultaneously hold coupons for more than one brand and 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 we consider coupons that may be valid for more than one period. In Section 5 we analyze a real data set of purchases for which we estimate consumer preferences with and without knowledge of the individual purchases and coupon usage. In Section 6 we relate our empirical results to some of the findings in the literature about the coordination of coupon promotions and pricing and we also show how these results can be used to analyze important managerial questions. Finally, in Section 7 we conclude this article discussing 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 the market. In each period $t$, $N_{jt}^c \leq N$ consumers receive one coupon for brand $j$ and $N_{jt}^r \leq N_{jt}^c$ of them redeem their coupons. We specify the following assumptions regarding information availability and the distribution, consideration and redemption of coupons:

**(A1) Single Coupon.** Each consumer may have a coupon for at most 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.

**(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.

**(A6) Coupon Consideration.** The remaining coupons for non-purchased brands (non-redeemed coupons) were considered by consumers in their purchase decisions, but just not used.

In addition to these assumptions about coupon usage, we assume (as is standard) that consumers choose the product with the highest utility and the choice of consumer $i$ in period $t$ ($y_{it}$) satisfies:

$$
\begin{aligned}
y_{it} &= \text{argmax}_j \;\; U_{ijt} \\
&= \text{argmax}_j \;\; V_{ijt} + \epsilon_{ijt} \\
&= \text{argmax}_j \;\; \phi_i' x_{jt} + \psi_i c_{ijt} + \epsilon_{ijt},
\end{aligned} \tag{1}
$$

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., $V_{ijt} = \phi_{it}' X_{jt} + \psi_i c_{ijt}$); $X_{jt}$ 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; $\phi_i$ and $\psi_i$ are utility coefficients for consumer $i$, where the latter measures the utility trade off for consumer $i$ of using a coupon[1]; and, $\epsilon_{ijt}$ is an individual-specific demand shock for the utility of alternative $j$ for consumer $i$ in period $t$.

Assuming $\epsilon_{ijt}$ 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, 1985):

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

[1]One possible extension of the model is to allow for different coupon coefficients for different brands (i.e., replace $\psi_i$ by $\psi_{ij}$). 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 modelled by the researcher, this may also reduce the degrees of freedom for estimating the response of consumers to coupon promotions.

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 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:

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

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

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

$$\sum_{j=1}^{J} c_{ijt} \leq 1 \qquad \text{(Maximum Number of Coupons)}, \tag{6}$$

where these restrictions are related to the market share (equation 3), the total number of distributed coupons (equation 4), the total number of redemptions (equation 5), and the maximum number of coupons that a consumer may simultaneously hold (equation 6), respectively, for any brand $j$ in each period $t$.

Finally, we model the heterogeneity in consumer preferences 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 $\overline{\theta}$ and variance-covariance matrix $D$, as is common in empirical applications (e.g., Berry, Levinsohn and Pakes 1995).

## 2.1 Likelihood Function and Posterior Density

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

$$\mathcal{L}_{aug} = \left( \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{t=1}^{T} p_{ijt}(c_{ijt})^{z_{ijt}} \right) \mathrm{I}_{\{(Z,C)\in\Omega\}}, \tag{7}$$

where:

$$\Omega = \left\{ (Z,C) : \sum_{i=1}^{N} z_{ijt} = NS_{jt}, \sum_{i=1}^{N} c_{ijt} = N_{jt}^{c}, \sum_{i=1}^{N} c_{ijt} z_{ijt} = N_{jt}^{r}, \sum_{j=1}^{J} c_{ijt} \leq 1 \right\}; \tag{8}$$

the indicator function ensures that the (augmented) individual choices and coupon variables are exactly consistent with the aggregate information according to equations (3), (4), (5) and (6); and $\Omega$ is set of all possible configurations of choices and coupon usage $(Z, C)$ satisfying these constraints. We 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 (7), the posterior density of the parameters and the augmented data $(Z, C)$ is proportional to the following expression:

$$f(Z, C, \theta, \overline{\theta}, D | S, N^{c}, N^{r}, X) \propto \left( \prod_{i=1}^{N} N(\theta_i | \; \overline{\theta}, D) \prod_{j=1}^{J} \prod_{t=1}^{T} p_{ijt}(c_{it})^{z_{ijt}} \right) \mathrm{I}_{\{(Z,C)\in\Omega\}} \pi(\overline{\theta}, D), \tag{9}$$

where $\theta$ is a matrix containing each of the vectors of individual coefficients ($\theta_i$); $S$ 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; $N(\cdot|\overline{\theta}, D)$ is the density of a multivariate random variable with mean $\overline{\theta}$ and variance-covariance matrix $D$; and $\pi(\overline{\theta}, D)$ is the hyperprior for $\overline{\theta}$ and $D$, which is specified here as a standard Normal-Inverted Wishart prior (see Gelman, Carlin, Stern and Rubin, 1995, p. 80).

After formulating the augmented likelihood and the posterior density, we discuss in the next subsection 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, we generate draws from their full-conditional posterior distribution (Gibbs sampling) and we use these values to make posterior inferences about these parameters and, hence, inferences for assessing the effectiveness of coupons.

## 2.2 Estimation

As in Chen and Yang (2004) and Musalem, Bradlow and Raju (2004), 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 $\{\theta_i\}_{i=1}^{R}$, $\overline{\theta}$ and $D$ can be sampled using standard Bayesian methods (Allenby and Rossi 2003); for brevity not described here. Therefore, we focus 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 (2004) which only considered one type of restriction (market share). Under the generalization presented here, the augmented individual choices and coupons for each period, $(z_t, c_t)$, will be drawn directly and jointly from their full-conditional posterior distribution (Gibbs sampling) and these must satisfy the constraints defined in equations (3), (4), (5) and (6). For computational convenience, we propose sampling $(z_t, c_t)$ by first assigning consumers to pairs and then sequentially updating the choices and coupons in each pair. Specifically, suppose we consider the choices and coupons of consumers $i_1$ and $i_2$ in period $t$, conditioning on all other parameters including the choices and coupons of all other consumers. Then, using (9), the full-conditional posterior distribution of the choices and coupons of these two consumers in period $t$ is given by:

$$f(z_{i_1t}, c_{i_1t}, z_{i_2t}, c_{i_2t}|*) = K \cdot \mathrm{I}_{\{(Z,C)\in\Omega\}} \prod_{j=1}^{J} p_{i_1jt}(c_{i_1t})^{z_{i_1jt}} p_{i_2jt}(c_{i_2t})^{z_{i_2jt}} \tag{10}$$

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 Markov Chain, the values of $Z$ and $C$ satisfy constraints (3), (4), (5) and (6), it is easy to verify that when the choices and coupons of all other consumers are held constant, there are only two instances of $\{(z_{i_1t}, c_{i_1t}), (z_{i_2t}, c_{i_2t})\}$ that have a non-zero probability. The first corresponds to the current values of $\{(z_{i_1t}, c_{i_1t}), (z_{i_2t}, c_{i_2t})\}$, while in the second instance consumers $i_1$ and $i_2$ interchange their choices and coupons in period $t$. Note that any other configuration would violate one or more of the constraints (3), (4), (5) and (6). Accordingly, the full-conditional posterior probability of the event where the choices and coupons of these two consumers take their current values corresponds to:

$$f(z_{i_1t}, c_{i_1t}, z_{i_2t}, c_{i_2t}|*) = \frac{\prod_{j=1}^{J} p_{i_1jt}(c_{i_1t})^{z_{i_1jt}} p_{i_2jt}(c_{i_2t})^{z_{i_2jt}}}{\prod_{j=1}^{J} p_{i_1jt}(c_{i_1t})^{z_{i_1jt}} p_{i_2jt}(c_{i_2t})^{z_{i_2jt}} + \prod_{j=1}^{J} p_{i_1jt}(c_{i_2t})^{z_{i_2jt}} p_{i_2jt}(c_{i_1t})^{z_{i_1jt}}}, \tag{11}$$

while the complement of this expression defines the probability of interchanging the choices and coupons of both consumers. Based on this result, the details of the procedure for simulating $Z$ and $C$ are described in Appendix A. In the next subsection, we illustrate this method using a numerical experiment.

## 2.3 Numerical Experiment

In order to demonstrate the efficacy of this approach, we consider a small numerical example with $J = 3$ brands, $T = 50$ periods and $N = 500$ consumers. 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 the coupon indicator ($c_{ijt}$). The true mean and variance of the individual coefficients ($\theta_i$) were chosen as $\overline{\theta}$ =(1 1 -1 1)$'$ and $D = I_4$, respectively, where $I_4$ denotes the identity matrix with four rows and columns. In addition, coupons were randomly assigned to the simulated consumers, where the probability of a receiving a coupon in a given period was chosen to be equal to 0.3 for all consumers in all periods (i.e., $P(\sum_{j=1}^{3} c_{ijt} = 1) = 0.3$). Finally, each of these coupons is only valid for one purchase of a single brand, where the corresponding brand is randomly selected with probability 1/3 (i.e., $P(c_{ijt} = 1 \mid \sum_{j=1}^{3} c_{ijt} = 1) = 1/3$, for all $j$).

Using only aggregate information (i.e., market shares, number of redeemed coupons and number of coupons available for each brand in each period), we estimated $\overline{\theta}$ and $D$ according to the procedure described in this section. The starting values for $\overline{\theta}$ and $D$ correspond to $\overline{\theta}$ = (0 0 0 0)' and $D = 0.1I_4$. The starting values for our 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), (5) and (6). In addition, the following hyperprior distributions are specified: $\overline{\theta} \sim \mathcal{N}(0, 10^5)$ and $D \sim$ Inverted 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 $\overline{\theta}$ and $D$, the mean and variance of the preference coefficients[2].

[2]Extensive simulation suggests that this was sufficient for convergence.

== Insert Table 1 here ==

From the results in Table 1 we observe that the true values of $\overline{\theta}$ 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. In the next section, we discuss the first two generalizations of the scenario analyzed here.

# 3 Multiple Coupons and Limited Information

In this section we introduce two extensions to the method described in §2. These extensions are related to our assumptions about coupon usage and the information that the researcher needs to collect in order to estimate the distribution of consumer preferences.

## 3.1 Multiple Coupons

In the previous section, we considered a situation where consumers may have at most one coupon. In this section, we now allow consumers to simultaneously hold coupons for more than one brand, while keeping all other assumptions from §2. This is a more realistic scenario given that sometimes different manufacturers promote their brands simultaneously and it is likely that a consumer might have access to coupons from more than one brand in a given period. Accordingly, we replace **A1** by:

**(A1') Multiple Coupons.** Each consumer may have coupons for more than one brand in each period.

In this case, it is no longer necessary to impose the constraint specified in (6). In terms of the estimation procedure, the absence of this constraint makes the joint sampling of choices and coupons from their posterior distribution more difficult. In the case analyzed in the previous section where we considered the choices and coupons of a pair of consumers, there were only two possible values that satisfy all the constraints: leave the choices and coupons at their current values or interchange them. In the case of multiple coupons, however, the number of feasible instances of $z$ and $c$ for a pair of consumers exponentially increases with the number of brands $J$, which makes the computation of the full-conditional posterior probability more cumbersome. As it will be shown later, 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.

## 3.2 Limited Information

In Section 2, we also assumed that the researcher had access to data on the number of redeemed coupons ($N^r_{jt}$) and the number of consumers that received a coupon ($N^c_{jt}$). The latter information on received coupons is much harder to get in practice and hence the extension presented here, which does *not* rely on this knowledge, is highly relevant for practical applications. In contrast, one could try to estimate $N^c_{jt}$ 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 to be different from the original number of distributed coupons. For example, some coupons might never reach any consumer, some consumers may lose their coupons, or some consumers may realize that they could have used a coupon when the coupon has already expired.

Accordingly, we generalize the estimation procedure by only requiring knowledge of the number of redeemed coupons, while the number of consumers that a received a coupon will be estimated according to the methodology that we present below. Consequently, we replace **A2** 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 $\delta_{jt}$ 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:

$$N_{jt}^r \leq \sum_{i=1}^{N} c_{ijt} \leq N\delta_{jt}. \tag{12}$$

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 $\delta_{jt}$. 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 $\delta_{jt}$ 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:

$$r_{jt} = P(c_{ijt} = 1) = \delta_{jt} \, \frac{e^{\alpha_j + \xi_{jt}}}{1 + e^{\alpha_j + \xi_{jt}}} \tag{13}$$

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

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

Finally, it is important to mention that the random effects for different brands $(\xi_{jt}, \xi_{j't})$ are allowed to be correlated via $\Sigma$. 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).

## 3.3 Estimation

We first describe in this subsection the updating of the unobserved individual choices and then we discuss how the updating of the coupon variables can be implemented.

### 3.3.1 Drawing choices from their full-conditional posterior distribution

As before, we start by considering pairs of consumers. In particular, suppose that we consider the choices of consumers $i_1$ and $i_2$ in period $t$ $(z_{i_1t}, z_{i_2t})$ conditioning on the coupon variables $(C)$ and all other parameters and assume that constraints (3), (4) and (5) are satisfied by the current values of $Z$ and $C$. Then, it follows that the only instances of $(z_{i_1t}, z_{i_2t})$ that satisfy the market share constraint (equation (3)) are the current values and the instance where these values are interchanged. In addition, it is necessary to take into account that a change in the choices of these two consumers $(z_{i_1t}, z_{i_2t})$ 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 $z_{i_2 t}$. The details of this procedure are presented in Appendix B.

### 3.3.2 Drawing coupons from their full-conditional posterior distribution

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 we have replaced the equality constraint specified in (4) by inequality (12), 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, we propose a Metropolis-Hastings (MH) step 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 (12). The details of this procedure are also presented in Appendix B.

### 3.3.3 Drawing $\delta$ from its full-conditional posterior distribution

As we mentioned before, when the observed number of redeemed coupons for brand $j$ in period $t$ is greater than zero, then by condition (12), $\delta_{jt}$ must be equal to 1. In addition, if in a given iteration $k$ there is a positive number of coupons assigned to brand $j$ in period $t$ (i.e., if $\sum_{i=1}^{N} c_{ijt} > 0$), then $\delta_{jt}$ must also be equal to 1 (see condition (12)). However, when no coupons are assigned, $\delta_{jt}$ must be imputed and we generate samples from its full-conditional posterior distribution by assigning $\delta_{jt}^{(k+1)} = 1$ with the following probability:

$$P(\delta_{jt}^{(k+1)} = 1 \mid * \,) = \frac{\left(\frac{1}{1+e^{\alpha_j+\xi_{jt}}}\right)^N \; q_j}{\left(\frac{1}{1+e^{\alpha_j+\xi_{jt}}}\right)^N \; q_j + (1-q_j)} \tag{14}$$

otherwise, $\delta_{jt}^{(k+1)} = 0$; where $\left(\frac{1}{1+e^{\alpha_j+\xi_{jt}}}\right)^N$ is the probability that none of the $N$ consumers received one of the coupons distributed for brand $j$ in period $t$.

## 3.4 Numerical Experiment

We construct a numerical example based on the same parameter values for $\overline{\theta}$ and $D$ as in §2.3. The true values for the parameters that determine the coupon probabilities are specified as: $q = (0.25, 0.50, 0.75)$, $\alpha = (-1, -3, -5)$ and

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

We specify a Beta(1,1) hyperprior distribution (i.e, Uniform$(0,1)$) for $q_1$, $q_2$ and $q_3$ (i.e., $a_j = b_j = 1$), a $\mathcal{N}(0, 1000)$ for each $\alpha_j$ (i.e., $\sigma_j^2 = 1000$) and an Inverted Wishart$(5, 5I_3)$, weakly informative for $\Sigma$. As we mentioned before, we only use aggregate data on market shares and number of redeemed coupons for each brand in each period 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 $\overline{\theta}$, $D$ and $Z$ are the same as those used in §2.3, while the initial values for $\alpha_j$ and $\Sigma$ correspond to $\alpha_j = 0$ and $\Sigma_\xi = I_3$. In the case of the coupon variables ($C$), we first set an initial value for $N_{jt}^c$ equal to the integer part of $N_{jt}^r + 0.3(N - N_{jt}^r)$, and we then randomly assign these $N_{jt}^c$ coupons among the $N$ consumers. Using the method proposed in §3.3 we obtained the results presented in Table 2, 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 we observe that the true values of $\overline{\theta}$, $D$, $\alpha$, $q$ and $\Sigma$ are covered by their 95% posterior-probability intervals and that the posterior means and posterior medians are very close to the true values (within 1 posterior standard deviation).

# 4 Non-Immediate Expiration

In the preceding sections, we assumed that coupons *immediately* expire after one period. If the length of a period is equal to a week or less, 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'**, while keeping all other assumptions from the previous section (i.e., **A1'**, **A2'**, **A4**, **A5**, **A6**):

**(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 expire 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:

$$(15) \qquad r_{ijt}(c_{ijt-1}, z_{ijt-1}) = P(c_{ijt} = 1) = \delta_{jt} \, \frac{e^{\alpha_j + \alpha_{J+1} c_{ijt-1}(1-z_{ijt-1}) + \xi_{jt}}}{1 + e^{\alpha_j + \alpha_{J+1} c_{ijt-1}(1-z_{ijt-1}) + \xi_{jt}}}, \qquad 2 \leq t < T,$$

where $\alpha_{J+1}$ measures the change in the coupon-availability probability that is triggered when a consumer had a coupon in the previous period which was not redeemed. For example, positive values of $\alpha_{J+1}$ imply that if a coupon was available to consumer $i$ in period $t-1$, but the consumer did not use it (i.e., if $c_{ijt-1}(1 - z_{ijt-1}) = 1$), then the probability that the consumer will have a valid coupon in period $t$ is higher.

In addition, it is also necessary to define coupon probabilities for the first period that do not depend on $c_{ij0}$ or $z_{ij0}$, data usually unobservable. Hence, we instead specify a different model for $r_{ij1}$, which is defined as follows[3]:

$$r_{ij1} = P(c_{ij1} = 1) = \delta_{j1} \, \frac{e^{\alpha_j + \gamma_j}}{1 + e^{\alpha_j + \gamma_j}}. \tag{16}$$

As in the previous section, we use a Normal hyperprior distribution for each of the components of $\alpha$ and $\gamma$: $\alpha_l \sim \mathcal{N}(0, \sigma^2_{\alpha_l})$, $l = 1, .., J+1$, $\gamma_j \sim \mathcal{N}(0, \sigma^2_{\gamma_j})$, $j = 1, .., J$. We also note that $\gamma_j$ 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 $\gamma_j$. If that probability is very small, then we can simply ignore $\gamma_j$ for any practical purposes.

Finally, it is important to mention that the specification of the coupon probabilities in equation (15) introduces a correlation between choices and coupons of consecutive periods (unless $\alpha_{J+1} = 0$). Accordingly, we must redefine the steps that are necessary to sample choices and coupons $(Z, C)$ from their full-conditional posterior distribution in order to capture this autocorrelation in the (unobserved) time series of coupons and choices. These modifications are explained in Appendix C.

## 4.1 Numerical Experiment

We constructed a numerical example based on the same parameter values for $\overline{\theta}$, $D$, $q$ and $\Sigma$ as in §3.4. The true values for the new additional parameters are $\alpha = (-2, -1, 0, 2)$ and

[3] We 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.

$\gamma = (1, 0, -1)$. In addition, we constrain $\delta_{j1}$ to be equal to 1 for every brand (i.e., all brands distributed coupons in period 1) in order to get meaningful estimation results for each of the components of $\gamma$ (in our real data analyses in §5 this is not required).

We use the same hyperprior distributions and starting values for $\overline{\theta}$, $D$, $q$, $\alpha$ and $\Sigma$ as in §3.4. The prior for each $\gamma_j$ corresponds to $N(0, 10)$ (i.e., $\sigma^2_{\gamma_j} = 10$, $j = 1, ..., J$) and the starting value corresponds to $\gamma = (0, 0, 0)$ . Using the method described in Appendix C, we obtained the results presented in Table 3, where there results are based on a single run of 400,000 iterations with the last 200,000 used for estimation. From the results we observe that the true values of $\overline{\theta}$, $D$, $q$, $\alpha$, $\gamma$ and $\Sigma$ are covered by their 95% posterior-probability intervals and, in most cases (25 out of 30 parameters) the posterior means are within 1 posterior standard deviation from the true values, but none deviate much.

== Insert Table 3 here ==

In summary, our simulations demonstrate the efficacy of our approach under the most basic to more realistic conditions. We now apply these methods to a real data set.

# 5 Empirical Application

In this section, we apply the methods described in Sections 3 and 4 to a data set of purchases in the ice cream product category. In order to provide an empirical validation of these methods, we use a data set for which disaggregate data are available and we implement two separate estimation procedures: disaggregate and aggregate estimation. While normally this would not be available, fitting both methods and comparing their results provides an empirical test to validate our methodology.

In the first case, we use disaggregate data on choices and coupon redemption for eight different ice cream brands (Baldwin, Breyers, Country Charm, Deans Food, Dreyers Edys, Fieldcrest, Private Label, Sealtest), which were generated by a panel of consumers at a single

Table 5
Estimated Log-Marginal Likelihood

| Model | Disaggregate Estimation | Aggregate Estimation |
|---|---|---|
| Immediate Expiration | -15,105.956 | -16,263.287 |
| Non-Immediate Expiration | -14,352.553 | -12,573.062 |

store in an urban market in the period between June 1992 and June 1994. We selected a total of 165 panelists that made at least four purchases during the 99 weeks of observation[4]. In the second case, we only use the aggregate total number of choices and coupons redeemed for each brand in each week. Finally, in both cases we specify a utility function that includes a dummy variable for each brand ($x_1, ..., x_8$), prices ($x_9$) and feature ($x_{10}$). In addition, we include a non-purchase option in order to capture category expansion effects. Table 4 presents summary statistics for this data set.

Using the methods presented in Sections 3 (immediate expiration) and 4 (non-immediate expiration), we first assessed the degree of generality needed in the model. For model comparison purposes, we computed the log-marginal likelihood of the data (presented in Table 5) for each model (immediate and non-immediate expiration) under both estimation procedures (aggregate and disaggregate)[5]. From these results and according to the criterion in Kass and Raftery (1995) we find very strong empirical support for the second model (non-immediate expiration) under both estimation procedures (aggregate and disaggregate). In addition, we note that the fact that $\alpha_9$ (the coefficient for coupon carryover) is estimated to be significantly different from zero (see Table 6b) also provides support for selecting the second model instead of the first model.

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

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

In what follows, we focus the rest of our discussion on the results obtained for the selected model (non-immediate expiration), which are reported in Table 6. In the case of $\gamma$ and the off-diagonal elements of $D$ and $\Sigma$, we only report results for those elements that are estimated to be significantly different from zero for at least one of the estimation procedures. According to these results, we verified that the estimated 95% posterior probability intervals for each parameter overlap each other under both estimations for all parameters (except for $D_{66}$, $D_{67}$, $D_{69}$). In terms of the demand parameters (see Table 6a), the estimated posterior means for $\overline{\theta}$ under both cases are fairly close (in most cases, 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 and price 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. In addition, we observe that the posterior standard deviations for $\overline{\theta}$ 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 4, 6 and 7 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 7 shows the estimated posterior mean (first block of results), 2.5%-ile (second block) and 97.5%-ile (third block) for each of these elasticities. These elasticities were computed assuming mean levels of prices, feature and coupon availability. These results show that the two sets of elasticities are reasonably close to each other and, therefore, both of them would generate similar managerial recommendations for pricing purposes, suggesting that our aggregate results well mimic the disaggregate ones.

In terms of the parameters related to the coupon probabilities ($\alpha$, $\gamma$, $\Sigma$ and $q$), we also observe a great degree of agreement between the two sets of estimated values (see Table 6b). In fact, the posterior means for each parameter under both estimations are within one posterior standard deviation from each other (except for $\alpha_3$).

Finally, we performed a series of posterior predictive checks (Gelman, Meng and Stern 1996) of the results obtained from the aggregate estimation, the strongest possible check of our approach. In particular, we computed the following check statistics from the imputed $(Z, C)$ under the aggregate estimation and compare them 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, we computed the corresponding true value using the disaggregate data and then we compared these true values with those estimated under the aggregate procedure. The results are presented in Figure 1 where $k$ is represented on the horizontal axis, the solid line represents the true values and the other three lines represent the 2.5%, 50.0% and 97.5% posterior quantiles. From these posterior predictive checks we observe that the true values for total purchases, coupon redemption and coupon penetration are, in general, within their 95% posterior-probability intervals. For the other two measures, penetration and number of different brands, the estimated values for 5 of the 8 levels are within their 95% posterior probability intervals. A discussion of some extensions to our models that could potentially improve our results is presented in Section 7.

== Insert Figure 1 here ==

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

# 6 Discussion of Results and Managerial Implications

In this section, we relate the empirical results from our aggregate estimation to some of the findings in the literature about the coordination of coupon promotions and pricing. In addition, we also show how these results can be used to evaluate the impact of coupon promotion strategies on brand switching and purchase incidence.

We note that it is also possible to conduct a series of other interesting analyses using our empirical results from the aggregate estimation (results are available from the authors upon request). For example, we can compare the price and coupon sensitivity of coupon users and non-users within this product category and we can estimate what is the value of using consumer preference information for determining whether to distribute a coupon to a consumer. In particular, this last application would make targeting methods, such as those proposed by Rossi, McCulloch and Allenby (1996), even more powerful given that one could estimate the distribution of consumer preferences from aggregate data and then use data on a single choice (e.g., the current transaction of the customer) to decide whether to give a coupon to a consumer.

## 6.1 Prices and Coupon Promotions

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 get 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 the context of the empirical application presented in this paper, we can test whether prices are higher or lower when coupons are being offered. A simple test is to compute the ratio between the mean price for a given brand in periods $t$ where $\delta_{jt} = 1$ and the corresponding mean price for periods $t'$ where $\delta_{jt'} = 0$. The results of this estimation are reported in Table 8. Accordingly, we verified that the prices for the eight brands under analysis are estimated to be on average *lower* when coupons are being offered, where the estimated mean of this price ratio ranges between 0.95 and 0.98, however it is not significantly different from 1 for any of the eight brands (results are available upon request). Therefore, we conclude that coupons are not being used with the objective of charging a higher price to non-users while still getting some coupon users to buy the product at a reduced price. These results are similar to the findings of Anderson and Song (2004) across eight different product categories and Nevo and Wolfram (2002) in the breakfast cereal product category, although both papers offer different explanations for these results.

== Insert Table 8 here ==

### 6.2 Purchase Incidence and Brand Switching

An important issue from a managerial point of view is to analyze some of the behavioral consequences of coupons. In this respect, it is possible to decompose the incremental sales obtained by a brand into its purchase incidence and brand-switching components. The first corresponds to customers who switch from the no-purchase option, while the second considers those customers who switch from other brands. For illustrative purposes, we focus on brand 4 and we consider a period with mean levels of prices and feature and assume that coupons for all other brands are not available. Accordingly, we compute the incremental share of brand 4 for different levels of coupon distribution, ranging from 0% to 100% of consumers receiving a coupon for that brand. The decomposition of the incremental sales into its purchase incidence and brand switching components is shown in Figure 2.

From these results, we observe that a large proportion of the incremental sales corresponds to changes in purchase incidence (91.1% at the maximum incremental sales level). Therefore, coupons are primarily driving consumers to increase their consumption (at least in the short term), rather than changing their preferences among brands within this product category.

## 7 Conclusions

In this article, we have presented new methods for the estimation of demand models using 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 simulate the consequences of alternative coupon promotion strategies (policy simulations).

In terms of the estimation procedure developed in this paper, the basic idea is to simulate (data augment) 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 to some extent 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, while the coupon model could be modified in order to explicitly model the decision to redeem a coupon as suggested by Chiang (1995). 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); allowing for different preference coefficients for users and non-users of coupons (e.g., Bell and Chiang 2001); accounting for price-endogeneity in the estimation of the demand model (e.g., Yang, Chen and Allenby 2003); examining competitive effects (e.g., Besanko, Dubé and Gupta 2003); and exploring other assumptions about coupon availability, such as non-random systems (Manchanda, Rossi and Chintagunta 2004).

Finally, we believe that there are several other potential applications of these methods for dealing with aggregate data that may constitute valuable contributions to the marketing literature. In particular, other problems such as the study of customer retention, the evaluation of sales force performance and the response of consumers to stockouts may become interesting areas for the development of new methods for the estimation of models of individual behavior using aggregate or limited information.

Table 1
Results: Estimated posterior mean, standard deviation and quantiles for $\overline{\theta}$ and $D$ (single coupon).

| | $\overline{\theta}_1$ | $\overline{\theta}_2$ | $\overline{\theta}_3$ | $\overline{\theta}_4$ | $D_{11}$ | $D_{22}$ | $D_{33}$ | $D_{44}$ | $D_{12}$ | $D_{13}$ | $D_{14}$ | $D_{23}$ | $D_{24}$ | $D_{34}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 1.041 | 0.981 | -1.010 | 1.150 | 1.321 | 1.264 | 0.986 | 0.719 | 0.017 | 0.074 | 0.055 | 0.013 | 0.137 | -0.150 |
| std.dev. | 0.098 | 0.095 | 0.089 | 0.103 | 0.718 | 0.714 | 0.191 | 0.251 | 0.312 | 0.102 | 0.180 | 0.109 | 0.187 | 0.125 |
| 2.5% | 0.894 | 0.836 | -1.241 | 0.997 | 0.544 | 0.520 | 0.730 | 0.366 | -0.411 | -0.131 | -0.291 | -0.219 | -0.165 | -0.403 |
| 50.0% | 1.027 | 0.967 | -0.996 | 1.133 | 1.147 | 1.067 | 0.949 | 0.673 | -0.043 | 0.076 | 0.048 | 0.018 | 0.116 | -0.148 |
| 97.5% | 1.291 | 1.221 | -0.875 | 1.425 | 3.578 | 3.334 | 1.487 | 1.344 | 0.843 | 0.276 | 0.439 | 0.221 | 0.557 | 0.0888 |
| 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 |

Table 2

Results: Estimated posterior mean, standard deviation and quantiles for $\overline{\theta}$, $D$, $q$, $\alpha$ and $\Sigma$ (limited information).

| | $\overline{\theta}_1$ | $\overline{\theta}_2$ | $\overline{\theta}_3$ | $\overline{\theta}_4$ | $D_{11}$ | $D_{22}$ | $D_{33}$ | $D_{44}$ | $D_{12}$ | $D_{13}$ | $D_{14}$ | $D_{23}$ | $D_{24}$ | $D_{34}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

| | $q_1$ | $q_2$ | $q_3$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\Sigma_{11}$ | $\Sigma_{22}$ | $\Sigma_{33}$ | $\Sigma_{12}$ | $\Sigma_{13}$ | $\Sigma_{23}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 $\overline{\theta}$, $D$, $q$, $\alpha$, $\gamma$ and $\Sigma$ (non-immediate expiration).

| | $\overline{\theta}_1$ | $\overline{\theta}_2$ | $\overline{\theta}_3$ | $\overline{\theta}_4$ | $D_{11}$ | $D_{22}$ | $D_{33}$ | $D_{44}$ | $D_{12}$ | $D_{13}$ | $D_{14}$ | $D_{23}$ | $D_{24}$ | $D_{34}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 1.077 | 1.005 | -1.074 | 1.156 | 0.922 | 1.378 | 1.189 | 1.284 | 0.154 | 0.032 | -0.077 | -0.038 | 0.086 | 0.073 |
| std.dev. | 0.081 | 0.081 | 0.077 | 0.104 | 0.345 | 0.543 | 0.176 | 0.492 | 0.234 | 0.100 | 0.238 | 0.104 | 0.284 | 0.185 |
| 2.5% | 0.932 | 0.864 | -1.236 | 0.964 | 0.427 | 0.568 | 0.890 | 0.561 | -0.214 | -0.166 | -0.524 | -0.255 | -0.394 | -0.301 |
| 50.0% | 1.072 | 0.998 | -1.070 | 1.151 | 0.865 | 1.307 | 1.172 | 1.227 | 0.120 | 0.032 | -0.081 | -0.035 | 0.049 | 0.076 |
| 97.5% | 1.248 | 1.183 | -0.936 | 1.369 | 1.799 | 2.658 | 1.573 | 2.432 | 0.697 | 0.228 | 0.421 | 0.159 | 0.711 | 0.428 |
| 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 |

| | $q_1$ | $q_2$ | $q_3$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\gamma_1$ | $\gamma_2$ | $\gamma_3$ | $\Sigma_{11}$ | $\Sigma_{22}$ | $\Sigma_{33}$ | $\Sigma_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mean | 0.420 | 0.532 | 0.706 | -2.477 | -0.875 | -0.264 | 2.419 | 1.415 | -0.134 | -0.782 | 1.895 | 2.727 | 1.717 | 0.507 |
| std.dev. | 0.069 | 0.070 | 0.064 | 0.278 | 0.286 | 0.287 | 0.744 | 0.306 | 0.314 | 0.400 | 0.716 | 0.923 | 0.486 | 0.568 |
| 2.5% | 0.289 | 0.395 | 0.575 | -3.046 | -1.384 | -0.804 | 0.448 | 0.829 | -0.808 | -1.580 | 0.954 | 1.454 | 1.002 | -0.567 |
| 50.0% | 0.418 | 0.533 | 0.709 | -2.472 | -0.890 | -0.263 | 2.448 | 1.410 | -0.119 | -0.775 | 1.747 | 2.553 | 1.639 | 0.488 |
| 97.5% | 0.559 | 0.668 | 0.822 | -1.950 | -0.258 | 0.308 | 3.854 | 2.036 | 0.439 | -0.019 | 3.736 | 5.022 | 2.873 | 1.722 |
| True Values | 0.400 | 0.500 | 0.600 | -2.000 | -1.000 | 0.000 | 2.000 | 1.000 | 0.000 | -1.000 | 2.000 | 2.000 | 2.000 | 1.000 |

| | $\Sigma_{13}$ | $\Sigma_{23}$ |
|---|---|---|
| mean | 0.235 | -0.622 |
| std.dev. | 0.517 | 0.526 |
| 2.5% | -0.796 | -1.799 |
| 50.0% | 0.229 | -0.575 |
| 97.5% | 1.295 | 0.300 |
| True Values | 0.000 | -1.000 |

Table 4
Summary Statistics for the ice cream data.

| Variable | Mean | Std. Dev. | Min. | Max. | Obs. |
|---|---|---|---|---|---|
| $S_1$ | 0.006 | 0.011 | 0.000 | 0.055 | T= 99 |
| $S_2$ | 0.021 | 0.021 | 0.000 | 0.115 | 99 |
| $S_3$ | 0.011 | 0.016 | 0.000 | 0.085 | 99 |
| $S_4$ | 0.005 | 0.009 | 0.000 | 0.042 | 99 |
| $S_5$ | 0.008 | 0.020 | 0.000 | 0.133 | 99 |
| $S_6$ | 0.025 | 0.019 | 0.000 | 0.133 | 99 |
| $S_7$ | 0.016 | 0.020 | 0.000 | 0.103 | 99 |
| $S_8$ | 0.021 | 0.017 | 0.000 | 0.079 | 99 |
| $x_{91}$ | 3.461 | 0.594 | 1.990 | 3.990 | T= 99 |
| $x_{92}$ | 3.162 | 0.410 | 1.690 | 3.490 | 99 |
| $x_{93}$ | 3.167 | 0.488 | 1.830 | 3.590 | 99 |
| $x_{94}$ | 3.152 | 0.489 | 1.740 | 3.590 | 99 |
| $x_{95}$ | 3.881 | 0.699 | 2.050 | 4.350 | 99 |
| $x_{96}$ | 1.609 | 0.141 | 0.850 | 1.690 | 99 |
| $x_{97}$ | 2.793 | 1.101 | 1.370 | 4.190 | 99 |
| $x_{98}$ | 2.580 | 0.180 | 1.990 | 2.890 | 99 |
| $x_{101}$ | 0.171 | 0.378 | 0.000 | 1.000 | T= 99 |
| $x_{102}$ | 0.271 | 0.446 | 0.000 | 1.000 | 99 |
| $x_{103}$ | 0.183 | 0.378 | 0.000 | 1.000 | 99 |
| $x_{104}$ | 0.171 | 0.373 | 0.000 | 1.000 | 99 |
| $x_{105}$ | 0.126 | 0.312 | 0.000 | 1.000 | 99 |
| $x_{106}$ | 0.107 | 0.305 | 0.000 | 1.000 | 99 |
| $x_{107}$ | 0.117 | 0.283 | 0.000 | 1.000 | 99 |
| $x_{108}$ | 0.212 | 0.411 | 0.000 | 1.000 | 99 |
| $N_1^r$ | 0.162 | 0.889 | 0.000 | 8.000 | T= 99 |
| $N_2^r$ | 0.242 | 0.959 | 0.000 | 8.000 | 99 |
| $N_3^r$ | 0.505 | 1.955 | 0.000 | 14.000 | 99 |
| $N_4^r$ | 0.162 | 0.792 | 0.000 | 7.000 | 99 |
| $N_5^r$ | 0.222 | 1.374 | 0.000 | 13.000 | 99 |
| $N_6^r$ | 0.192 | 1.811 | 0.000 | 18.000 | 99 |
| $N_7^r$ | 0.475 | 1.913 | 0.000 | 13.000 | 99 |
| $N_8^r$ | 0.444 | 1.263 | 0.000 | 10.000 | 99 |

Table 6a

Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for $\overline{\theta}$ and $D$ (non-immediate expiration, iterations 350,001-700,000).

| | Agg. mean | Agg. (s.d.) | Agg. 2.5% | Agg. 97.5% | Disagg. mean | Disagg. (s.d.) | Disagg. 2.5% | Disagg. 97.5% |
|---|---|---|---|---|---|---|---|---|
| $\overline{\theta}_1$ | -2.337 | (0.409) | -3.217 | -1.577 | -2.464 | (0.345) | -3.152 | -1.780 |
| $\overline{\theta}_2$ | -0.937 | (0.343) | -1.648 | -0.311 | -1.195 | (0.271) | -1.730 | -0.664 |
| $\overline{\theta}_3$ | -2.284 | (0.460) | -3.310 | -1.503 | -1.836 | (0.278) | -2.391 | -1.304 |
| $\overline{\theta}_4$ | -2.947 | (0.538) | -4.145 | -2.075 | -2.667 | (0.314) | -3.327 | -2.085 |
| $\overline{\theta}_5$ | -1.937 | (0.461) | -2.940 | -1.106 | -1.516 | (0.316) | -2.165 | -0.910 |
| $\overline{\theta}_6$ | -2.140 | (0.236) | -2.689 | -1.740 | -2.643 | (0.244) | -3.136 | -2.174 |
| $\overline{\theta}_7$ | -2.105 | (0.298) | -2.755 | -1.606 | -2.120 | (0.231) | -2.590 | -1.679 |
| $\overline{\theta}_8$ | -1.417 | (0.336) | -2.124 | -0.816 | -1.624 | (0.286) | -2.206 | -1.067 |
| $\overline{\theta}_9$ | -1.329 | (0.084) | -1.501 | -1.174 | -1.371 | (0.085) | -1.542 | -1.207 |
| $\overline{\theta}_{10}$ | 0.297 | (0.156) | -0.012 | 0.600 | 0.366 | (0.095) | 0.178 | 0.551 |
| $\overline{\theta}_{11}$ | 3.070 | (0.499) | 2.054 | 3.943 | 2.432 | (0.344) | 1.780 | 3.118 |
| $D_{11}$ | 2.024 | (0.880) | 0.836 | 4.386 | 2.795 | (1.016) | 1.309 | 5.120 |
| $D_{22}$ | 1.620 | (0.742) | 0.717 | 3.557 | 2.432 | (0.733) | 1.309 | 4.185 |
| $D_{33}$ | 2.347 | (1.189) | 0.870 | 5.581 | 2.535 | (0.803) | 1.250 | 4.357 |
| $D_{44}$ | 2.481 | (1.296) | 0.897 | 5.960 | 2.566 | (0.870) | 1.208 | 4.567 |
| $D_{55}$ | 3.403 | (1.437) | 1.279 | 6.946 | 2.687 | (1.043) | 1.183 | 5.168 |
| $D_{66}$ | 1.267 | (0.440) | 0.644 | 2.364 | 4.781 | (0.915) | 3.211 | 6.775 |
| $D_{77}$ | 1.754 | (0.718) | 0.798 | 3.512 | 3.141 | (0.736) | 1.898 | 4.753 |
| $D_{88}$ | 1.620 | (0.671) | 0.716 | 3.230 | 4.547 | (1.125) | 2.728 | 7.059 |
| $D_{99}$ | 0.251 | (0.051) | 0.173 | 0.369 | 0.371 | (0.084) | 0.242 | 0.570 |
| $D_{1010}$ | 1.368 | (0.433) | 0.716 | 2.407 | 0.514 | (0.101) | 0.346 | 0.738 |
| $D_{1111}$ | 3.402 | (1.659) | 1.209 | 7.670 | 2.229 | (0.655) | 1.221 | 3.757 |
| $D_{18}$ | 0.303 | (0.595) | -0.656 | 1.724 | 1.523 | (0.837) | 0.097 | 3.386 |
| $D_{34}$ | 0.072 | (0.815) | -1.622 | 1.831 | 1.532 | (0.726) | 0.339 | 3.198 |
| $D_{38}$ | 0.131 | (0.659) | -1.119 | 1.657 | 1.622 | (0.783) | 0.316 | 3.338 |
| $D_{39}$ | -0.026 | (0.139) | -0.347 | 0.222 | -0.395 | (0.213) | -0.878 | -0.064 |
| $D_{48}$ | 0.290 | (0.601) | -0.784 | 1.649 | 1.526 | (0.838) | 0.117 | 3.334 |
| $D_{57}$ | 0.603 | (0.722) | -0.720 | 2.186 | 1.270 | (0.722) | 0.113 | 2.931 |
| $D_{59}$ | -0.324 | (0.195) | -0.769 | -0.032 | -0.435 | (0.253) | -1.048 | -0.068 |
| $D_{511}$ | 0.843 | (1.243) | -1.626 | 3.543 | 1.212 | (0.590) | 0.239 | 2.477 |
| $D_{67}$ | -0.115 | (0.361) | -0.888 | 0.584 | 2.607 | (0.663) | 1.455 | 4.049 |
| $D_{68}$ | 0.052 | (0.343) | -0.647 | 0.733 | 1.843 | (0.731) | 0.559 | 3.406 |
| $D_{69}$ | -0.061 | (0.091) | -0.258 | 0.108 | -0.745 | (0.229) | -1.247 | -0.356 |
| $D_{79}$ | -0.139 | (0.123) | -0.417 | 0.070 | -0.530 | (0.212) | -1.010 | -0.185 |
| $D_{89}$ | -0.089 | (0.125) | -0.376 | 0.118 | -0.656 | (0.252) | -1.234 | -0.261 |
| $D_{910}$ | -0.170 | (0.097) | -0.388 | -0.003 | -0.008 | (0.057) | -0.119 | 0.107 |

Table 6b

Empirical Results: Estimated posterior mean, standard deviation, 2.5% and 97.5% quantiles for $\alpha$, $\gamma$, $\Sigma$ and $q$ (non-immediate expiration, iterations 350,001-700,000).

| | Agg. mean | Agg. (s.d.) | Agg. 2.5% | Agg. 97.5% | Disagg. mean | Disagg. (s.d.) | Disagg. 2.5% | Disagg. 97.5% |
|---|---|---|---|---|---|---|---|---|
| $\alpha_1$ | -6.388 | (1.795) | -10.484 | -3.288 | -6.858 | (2.158) | -11.217 | -3.077 |
| $\alpha_2$ | -6.944 | (0.956) | -9.020 | -5.158 | -7.160 | (1.195) | -9.757 | -5.045 |
| $\alpha_3$ | -7.574 | (1.246) | -9.994 | -5.045 | -6.260 | (1.087) | -8.357 | -4.220 |
| $\alpha_4$ | -7.821 | (1.188) | -10.443 | -5.467 | -6.701 | (1.159) | -9.031 | -4.293 |
| $\alpha_5$ | -6.664 | (1.443) | -9.595 | -4.157 | -5.747 | (1.246) | -8.289 | -3.284 |
| $\alpha_6$ | -6.908 | (3.109) | -13.791 | -1.687 | -5.618 | (3.318) | -13.432 | -0.292 |
| $\alpha_7$ | -5.279 | (1.639) | -8.645 | -2.392 | -5.552 | (2.671) | -11.255 | -1.550 |
| $\alpha_8$ | -5.872 | (0.747) | -7.385 | -4.455 | -5.633 | (0.685) | -7.007 | -4.356 |
| $\alpha_9$ | 8.267 | (1.529) | 5.491 | 11.789 | 7.375 | (1.712) | 4.427 | 11.011 |
| $\gamma_8$ | 2.906 | (0.927) | 1.068 | 4.665 | 3.231 | (0.919) | 1.424 | 5.037 |
| $\Sigma_{11}$ | 5.927 | (4.308) | 1.385 | 18.633 | 8.528 | (6.491) | 1.639 | 25.640 |
| $\Sigma_{22}$ | 4.605 | (2.298) | 1.577 | 10.505 | 6.008 | (3.364) | 1.844 | 14.591 |
| $\Sigma_{33}$ | 10.566 | (4.878) | 3.444 | 22.199 | 8.164 | (3.805) | 2.966 | 17.361 |
| $\Sigma_{44}$ | 8.264 | (4.310) | 2.794 | 19.596 | 6.921 | (3.233) | 2.361 | 14.769 |
| $\Sigma_{55}$ | 5.790 | (3.579) | 1.574 | 14.963 | 4.484 | (2.494) | 1.384 | 10.938 |
| $\Sigma_{66}$ | 10.534 | (9.842) | 1.697 | 38.974 | 8.865 | (6.811) | 1.589 | 29.027 |
| $\Sigma_{77}$ | 8.168 | (5.055) | 2.142 | 21.603 | 10.915 | (7.622) | 2.326 | 30.866 |
| $\Sigma_{88}$ | 2.905 | (1.375) | 1.136 | 6.359 | 3.106 | (1.439) | 1.201 | 6.656 |
| $\Sigma_{34}$ | 7.235 | (3.737) | 2.218 | 16.813 | 5.444 | (2.590) | 1.611 | 11.877 |
| $q_1$ | 0.466 | (0.223) | 0.119 | 0.933 | 0.585 | (0.241) | 0.143 | 0.979 |
| $q_2$ | 0.699 | (0.176) | 0.349 | 0.982 | 0.785 | (0.160) | 0.407 | 0.992 |
| $q_3$ | 0.785 | (0.161) | 0.414 | 0.992 | 0.662 | (0.192) | 0.304 | 0.980 |
| $q_4$ | 0.763 | (0.176) | 0.363 | 0.991 | 0.710 | (0.193) | 0.300 | 0.988 |
| $q_5$ | 0.531 | (0.234) | 0.152 | 0.968 | 0.459 | (0.242) | 0.108 | 0.960 |
| $q_6$ | 0.179 | (0.189) | 0.015 | 0.741 | 0.163 | (0.202) | 0.012 | 0.798 |
| $q_7$ | 0.280 | (0.162) | 0.092 | 0.750 | 0.343 | (0.220) | 0.090 | 0.916 |
| $q_8$ | 0.678 | (0.156) | 0.383 | 0.966 | 0.735 | (0.137) | 0.456 | 0.976 |

Table 7
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 (non-immediate expiration, aggregate and disaggregate estimation).

| | | Agg. | | | | | | | | Disagg. | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| mean | 1 | -3.470 | 0.075 | 0.038 | 0.022 | 0.019 | 0.049 | 0.030 | 0.069 | -3.008 | 0.052 | 0.026 | 0.016 | 0.021 | 0.032 | 0.024 | 0.057 |
| | 2 | 0.021 | -3.220 | 0.032 | 0.020 | 0.019 | 0.043 | 0.027 | 0.066 | 0.014 | -2.587 | 0.027 | 0.013 | 0.019 | 0.029 | 0.016 | 0.037 |
| | 3 | 0.022 | 0.066 | -3.180 | 0.016 | 0.019 | 0.042 | 0.031 | 0.058 | 0.016 | 0.060 | -3.118 | 0.026 | 0.021 | 0.034 | 0.024 | 0.069 |
| | 4 | 0.025 | 0.081 | 0.032 | -3.259 | 0.018 | 0.048 | 0.028 | 0.072 | 0.019 | 0.059 | 0.052 | -3.102 | 0.021 | 0.025 | 0.028 | 0.072 |
| | 5 | 0.020 | 0.073 | 0.035 | 0.017 | -4.266 | 0.043 | 0.035 | 0.061 | 0.019 | 0.061 | 0.031 | 0.015 | -3.580 | 0.027 | 0.033 | 0.036 |
| | 6 | 0.019 | 0.060 | 0.029 | 0.016 | 0.016 | -1.909 | 0.026 | 0.056 | 0.012 | 0.041 | 0.023 | 0.008 | 0.012 | -2.276 | 0.051 | 0.058 |
| | 7 | 0.018 | 0.056 | 0.032 | 0.015 | 0.019 | 0.040 | -2.875 | 0.050 | 0.016 | 0.037 | 0.026 | 0.015 | 0.024 | 0.084 | -3.040 | 0.043 |
| | 8 | 0.020 | 0.070 | 0.030 | 0.019 | 0.017 | 0.043 | 0.025 | -2.799 | 0.017 | 0.040 | 0.034 | 0.018 | 0.012 | 0.044 | 0.020 | -3.107 |
| 2.5% | 1 | -4.167 | 0.027 | 0.008 | 0.009 | 0.005 | 0.025 | 0.013 | 0.034 | -3.722 | 0.029 | 0.014 | 0.008 | 0.011 | 0.017 | 0.013 | 0.032 |
| | 2 | 0.007 | -3.720 | 0.014 | 0.009 | 0.007 | 0.027 | 0.012 | 0.035 | 0.008 | -3.163 | 0.016 | 0.007 | 0.011 | 0.019 | 0.009 | 0.025 |
| | 3 | 0.004 | 0.027 | -3.899 | 0.005 | 0.007 | 0.021 | 0.013 | 0.024 | 0.008 | 0.033 | -3.785 | 0.013 | 0.011 | 0.018 | 0.013 | 0.038 |
| | 4 | 0.009 | 0.034 | 0.011 | -3.926 | 0.006 | 0.026 | 0.012 | 0.034 | 0.009 | 0.032 | 0.029 | -3.794 | 0.011 | 0.012 | 0.016 | 0.040 |
| | 5 | 0.005 | 0.025 | 0.012 | 0.005 | -5.044 | 0.020 | 0.015 | 0.026 | 0.008 | 0.033 | 0.016 | 0.007 | -4.314 | 0.014 | 0.017 | 0.018 |
| | 6 | 0.010 | 0.038 | 0.015 | 0.009 | 0.008 | -2.214 | 0.014 | 0.034 | 0.006 | 0.027 | 0.012 | 0.004 | 0.007 | -2.583 | 0.035 | 0.039 |
| | 7 | 0.006 | 0.023 | 0.012 | 0.005 | 0.008 | 0.019 | -3.533 | 0.024 | 0.007 | 0.021 | 0.013 | 0.007 | 0.013 | 0.049 | -3.628 | 0.023 |
| | 8 | 0.010 | 0.038 | 0.013 | 0.009 | 0.008 | 0.026 | 0.014 | -3.292 | 0.009 | 0.027 | 0.020 | 0.010 | 0.007 | 0.030 | 0.012 | -3.576 |
| 97.5% | 1 | -2.567 | 0.139 | 0.081 | 0.046 | 0.041 | 0.080 | 0.058 | 0.120 | -2.126 | 0.086 | 0.050 | 0.031 | 0.038 | 0.057 | 0.044 | 0.095 |
| | 2 | 0.040 | -2.548 | 0.063 | 0.041 | 0.039 | 0.064 | 0.047 | 0.109 | 0.026 | -2.067 | 0.047 | 0.024 | 0.032 | 0.045 | 0.031 | 0.055 |
| | 3 | 0.047 | 0.127 | -2.078 | 0.035 | 0.040 | 0.074 | 0.066 | 0.110 | 0.031 | 0.099 | -2.227 | 0.048 | 0.039 | 0.070 | 0.047 | 0.123 |
| | 4 | 0.052 | 0.164 | 0.067 | -2.298 | 0.039 | 0.078 | 0.055 | 0.128 | 0.037 | 0.096 | 0.090 | -2.364 | 0.039 | 0.049 | 0.051 | 0.118 |
| | 5 | 0.046 | 0.137 | 0.077 | 0.038 | -2.809 | 0.075 | 0.070 | 0.113 | 0.037 | 0.101 | 0.060 | 0.033 | -2.563 | 0.046 | 0.068 | 0.061 |
| | 6 | 0.033 | 0.090 | 0.052 | 0.029 | 0.030 | -1.573 | 0.045 | 0.084 | 0.025 | 0.066 | 0.052 | 0.017 | 0.021 | -1.943 | 0.083 | 0.093 |
| | 7 | 0.037 | 0.099 | 0.072 | 0.032 | 0.040 | 0.067 | -1.769 | 0.088 | 0.031 | 0.065 | 0.051 | 0.030 | 0.043 | 0.127 | -2.090 | 0.082 |
| | 8 | 0.037 | 0.115 | 0.058 | 0.035 | 0.033 | 0.065 | 0.044 | -2.181 | 0.030 | 0.059 | 0.073 | 0.033 | 0.021 | 0.066 | 0.042 | -2.506 |

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 $j$.

Table 8

Price Factor: ratio between the mean price for each brand $j$ during periods $t$ when $\delta_{jt} = 1$ and the mean price during periods $t'$ when $\delta_{jt'} = 0$.

| Brand | Mean | Std. Dev. | 2.5% | 97.5% |
|---|---|---|---|---|
| 1 | 0.981 | 0.046 | 0.903 | 1.072 |
| 2 | 0.958 | 0.031 | 0.910 | 1.020 |
| 3 | 0.980 | 0.054 | 0.891 | 1.092 |
| 4 | 0.986 | 0.058 | 0.877 | 1.111 |
| 5 | 0.952 | 0.047 | 0.854 | 1.035 |
| 6 | 0.955 | 0.053 | 0.786 | 1.016 |
| 7 | 0.976 | 0.098 | 0.804 | 1.165 |
| 8 | 0.978 | 0.014 | 0.955 | 1.006 |

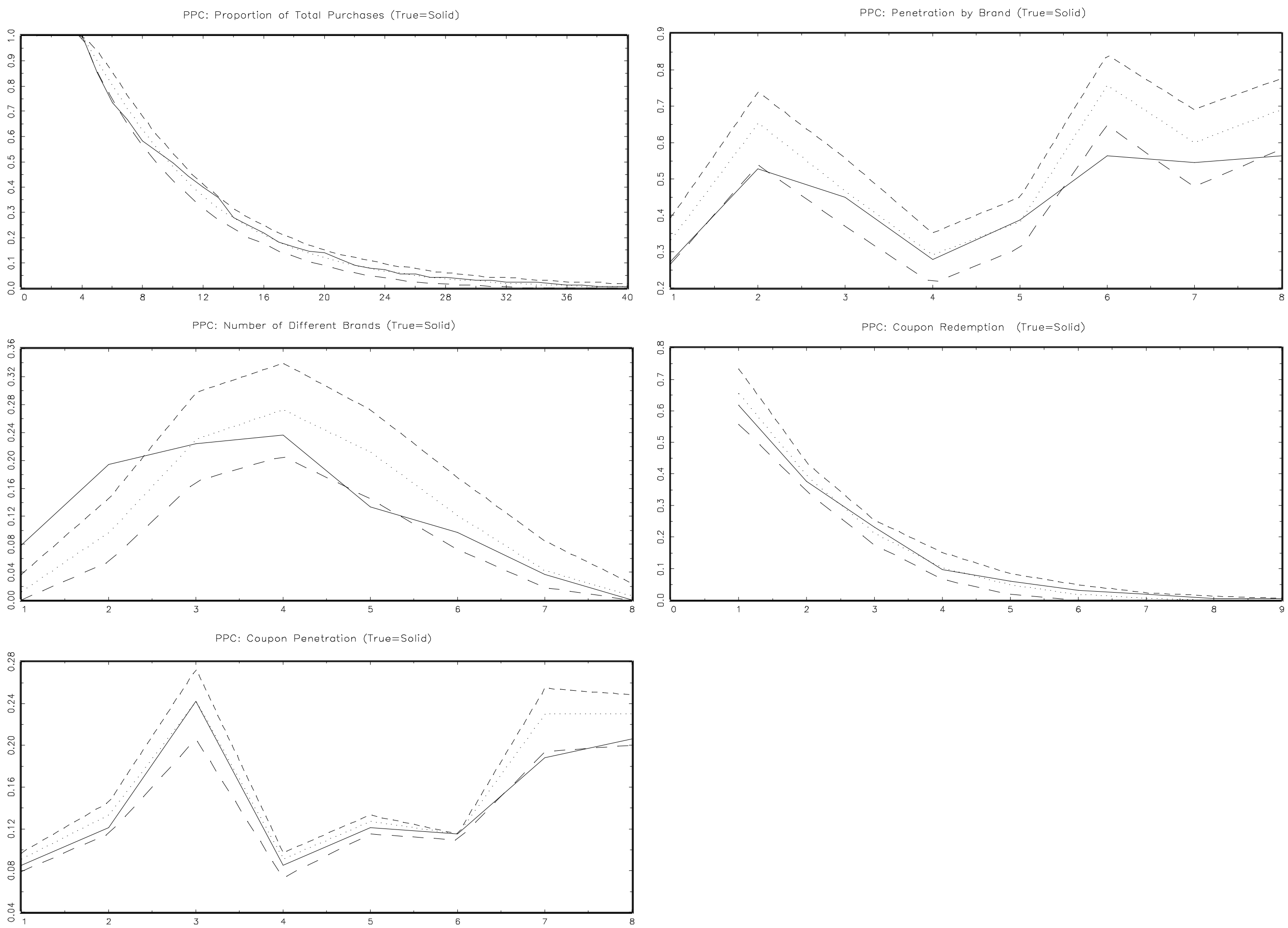


Figure 1: Posterior predictive checks.

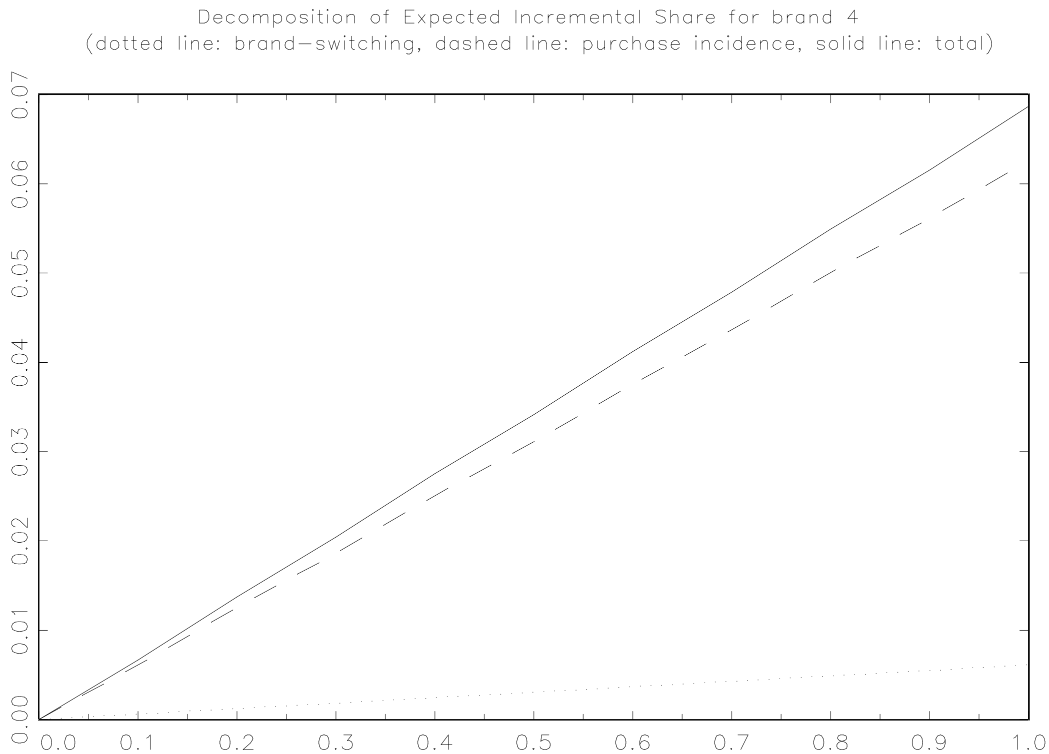


Figure 2: Decomposition of Incremental Share for Brand 4 vs. Coupon Distribution

## Appendix A: Sampling Choices and Coupons (single coupon case)

In this Appendix we describe the procedure to sample coupons and choices from their full-conditional posterior distribution according to the assumptions from Section 2.

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 $(z_{i_{1p}t}^{(k)}, c_{i_{1p}t}^{(k)})$ and $(z_{i_{2p}t}^{(k)}, c_{i_{2p}t}^{(k)})$ their choices and coupons in period $t$ in the current iteration $k$, respectively.

2. For each period $t$ and starting from the first pair, successively and jointly draw the choices and coupons $(z_{i_{1p}t}^{(k+1)}, c_{i_{1p}t}^{(k+1)})$ and $(z_{i_{2p}t}^{(k+1)}, c_{i_{2p}t}^{(k+1)})$ from their full-conditional posterior distribution. Dropping the pair $(p)$ and period $(t)$ subscripts for notational convenience, we proceed by assigning $(z_{i_1}^{(k+1)}, c_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}, c_{i_2}^{(k+1)}) = (z_{i_1}^{(k)}, c_{i_1}^{(k)}, z_{i_2}^{(k)}, c_{i_2}^{(k)})$ according to the following probability:

$$f\left((z_{i_1}^{(k+1)}, c_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}, c_{i_2}^{(k+1)}) = (z_{i_1}^{(k)}, c_{i_1}^{(k)}, z_{i_2}^{(k)}, c_{i_2}^{(k)}) \mid *\right) = \frac{\prod_{j=1}^{J} p_{i_1j}(c_{i_1}^{(k)})^{z_{i_1j}^{(k)}} p_{i_2j}(c_{i_2}^{(k)})^{z_{i_2j}^{(k)}}}{\prod_{j=1}^{J} p_{i_1j}(c_{i_1}^{(k)})^{z_{i_1j}^{(k)}} p_{i_2j}(c_{i_2}^{(k)})^{z_{i_2j}^{(k)}} + \prod_{j=1}^{J} p_{i_1j}(c_{i_2}^{(k)})^{z_{i_2j}^{(k)}} p_{i_2j}(c_{i_1}^{(k)})^{z_{i_1j}^{(k)}}} \tag{17}$$

   otherwise, exchange the choices and coupons of these two consumers by assigning: $(z_{i_1}^{(k+1)}, c_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}, c_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, c_{i_2}^{(k)}, z_{i_1}^{(k)}, c_{i_1}^{(k)})$.

## Appendix B: Sampling Choices and Coupons (multiple coupons and limited information)

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

### B.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:

$$f\left((z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, z_{i_1}^{(k)}) \mid *\right) = \frac{\prod_{j=1}^{J} p_{i_1j}^{z_{i_2j}^{(k)}} p_{i_2j}^{z_{i_1j}^{(k)}} \; \mathrm{I}_{\left\{c_{i_1j}z_{i_1j}^{(k)} + c_{i_2j}z_{i_2j}^{(k)} = c_{i_1j}z_{i_2j}^{(k)} + c_{i_2j}z_{i_1j}^{(k)}\right\}}}{\prod_{j=1}^{J} p_{i_1j}^{z_{i_2j}^{(k)}} p_{i_2j}^{z_{i_1j}^{(k)}} + \prod_{j=1}^{J} p_{i_1j}^{z_{i_1j}^{(k)}} p_{i_2j}^{z_{i_2j}^{(k)}}}, \tag{18}$$

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

Finally, we note that the full-conditional posterior probability in equation (18) can be rewritten as follows:

$$f\left((z_{i_1}^{(k+1)}, z_{i_2}^{(k+1)}) = (z_{i_2}^{(k)}, z_{i_1}^{(k)}) \mid *\right) = \frac{\mathrm{I}_{\left\{c_{i_1j}z_{i_1j}^{(k)}+c_{i_2j}z_{i_2j}^{(k)}=c_{i_1j}z_{i_2j}^{(k)}+c_{i_2j}z_{i_1j}^{(k)}\right\}}}{1+\dfrac{\prod_{j=1}^{J} p_{i_1j}^{z_{i_2j}^{(k)}} p_{i_2j}^{z_{i_1j}^{(k)}}}{\prod_{j=1}^{J} p_{i_1j}^{z_{i_1j}^{(k)}} p_{i_2j}^{z_{i_2j}^{(k)}}}}. \tag{19}$$

**B.2) Sampling Coupons:**

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 Kernel[6].
   (c) Accept $c_{it}^*$, according to the following MH probability that takes into account the likelihood of coupons and choices:

$$P\left(c_{it}^{(k+1)} = c_{it}^*\right) = \frac{\prod_{j=1}^{J} p_{ijt}(c_{it}^*)^{z_{ijt}} \ r_{jt}^{c_{ijt}^*}(1-r_{jt})^{1-c_{ijt}^*}}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}} \ r_{jt}^{c_{ijt}^{(k)}}(1-r_{jt})^{1-c_{ijt}^{(k)}}}, \tag{20}$$

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

## Appendix C: Sampling Choices and Coupons (non-immediate expiration)

In this Appendix we describe the procedure to sample coupons and choices from their full-conditional posterior distribution according to the assumptions from Section 4.

**C.1) Sampling Choices:**
As before, we consider the choices of a pair of consumers and we must decide whether to interchange their choices or leave them at their current values. In this case, we follow the same procedure

[6]One 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 ($c_{it}^*$).

described in B.1 but we replace the full-conditional posterior probability in equation (19) by the following expression:

$$(21) \qquad f\left((z_{i_1t}^{(k+1)}, z_{i_2t}^{(k+1)}) = (z_{i_2t}^{(k)}, z_{i_1t}^{(k)}) \mid *\right)$$

$$= \begin{cases} \dfrac{\mathrm{I}_{\left\{c_{i_1jt}z_{i_1jt}^{(k)}+c_{i_2jt}z_{i_2jt}^{(k)}=c_{i_1jt}z_{i_2jt}^{(k)}+c_{i_2jt}z_{i_1jt}^{(k)}\right\}}}{1+\dfrac{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_2jt}^{(k)}} p_{i_2jt}^{z_{i_1jt}^{(k)}} h\left(c_{i_1jt+1}\middle|c_{i_1jt},z_{i_2jt}^{(k)}\right) h\left(c_{i_2jt+1}\middle|c_{i_2jt},z_{i_1jt}^{(k)}\right)}{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_1jt}^{(k)}} p_{i_2jt}^{z_{i_2jt}^{(k)}} h\left(c_{i_1jt+1}\middle|c_{i_1jt},z_{i_1jt}^{(k)}\right) h\left(c_{i_2jt+1}\middle|c_{i_2jt},z_{i_2jt}^{(k)}\right)}}, & 1 \le t \le T-1; \\[2ex] \dfrac{\mathrm{I}_{\left\{c_{i_1jt}z_{i_1jt}^{(k)}+c_{i_2jt}z_{i_2jt}^{(k)}=c_{i_1jt}z_{i_2jt}^{(k)}+c_{i_2jt}z_{i_1jt}^{(k)}\right\}}}{1+\dfrac{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_2jt}^{(k)}} p_{i_2jt}^{z_{i_1jt}^{(k)}}}{\prod_{j=1}^{J} p_{i_1jt}^{z_{i_1jt}^{(k)}} p_{i_2jt}^{z_{i_2jt}^{(k)}}}}, & \text{t=T.} \end{cases}$$

where $h(\cdot|\cdot,\cdot)$ is the likelihood contribution of next-period coupons based on current coupons and choices. This function is defined as follows:

$$(22) \qquad h(c_{ijt+1}|\, c_{ijt}, z_{ijt}) = (r_{ijt+1}(c_{ijt}, z_{ijt}))^{c_{ijt+1}} \left(1 - r_{ijt+1}(c_{ijt}, z_{ijt})\right)^{1-c_{ijt+1}}, \; t = 1, .., T-1.$$

**C.2) Sampling Coupons:**

The updating of the coupon variables can be implemented following the same procedure described in Appendix B but replacing the MH probability in (20) by the following expression:

$$(23) \qquad P\left(c_{it}^{(k+1)} = c_{it}^{*}\right) = \begin{cases} \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^{*})^{z_{ijt}} \; h(c_{ijt+1}|c_{ijt}^{*},z_{ijt}) \; r_{ij1}^{c_{ijt}^{*}} (1-r_{ij1})^{(1-c_{ijt}^{*})}}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}} \; h(c_{ijt+1}|c_{ijt}^{(k)},z_{ijt}) \; r_{ij1}^{c_{ijt}^{(k)}} (1-r_{ij1})^{(1-c_{ijt}^{(k)})}}, & t = 1; \\[2ex] \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^{*})^{z_{ijt}} \; h(c_{ijt+1}|c_{ijt}^{*},z_{ijt}) \; h(c_{ijt}^{*}|c_{ijt-1},z_{ijt-1})}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}} \; h(c_{ijt+1}|c_{ijt}^{(k)},z_{ijt}) \; h(c_{ijt}^{(k)}|c_{ijt-1},z_{ijt-1})}, & 2 \le t \le T-1; \\[2ex] \dfrac{\prod_{j=1}^{J} p_{ijt}(c_{it}^{*})^{z_{ijt}} \; h(c_{ijt}^{*}|c_{ijt-1},z_{ijt-1})}{\prod_{j=1}^{J} p_{ijt}(c_{it}^{(k)})^{z_{ijt}} \; h(c_{ijt}^{(k)}|c_{ijt-1},z_{ijt-1})}, & t = T. \end{cases}$$

## Appendix D: Estimation of the marginal likelihood

In what follows we derive an estimator of the marginal likelihood by generalizing the harmonic mean method proposed by Newton and Raftery (1994). This generalization is needed for the aggregate estimation procedures presented in this paper that are based on augmenting the aggregate data $(A)$ with unobserved sequences of choices $(Z)$ and coupons $(C)$.

Let $\Omega_{\mathcal{M}}$ denote the set of all values of $(Z, C)$ consistent with the aggregate data $(A)$ under model $\mathcal{M}$ and let $\varphi$ denote the collection of parameters that determine the likelihood of the augmented choices and coupons (i.e., $\varphi = \{\theta, r\}$). We are interested in computing $p(A|\mathcal{M})$, the marginal likelihood of the aggregate data $A$ under model $\mathcal{M}$. For notational convenience, we drop the model subscript ($\mathcal{M}$) and we refer to $p(A|\mathcal{M})$ and $\Omega_{\mathcal{M}}$ simply as $p(A)$ and $\Omega$, respectively. By noting that $\int p(\varphi)d\varphi = 1$, it is straightforward to verify that the marginal likelihood $p(A)$ satisfies the following equation:

$$\frac{1}{p(A)} = \frac{1}{|\Omega|}\sum_{(Z,C)\in\Omega}\int \frac{p(\varphi)}{p(A)}d\varphi. \tag{24}$$

Using Bayes' rule and noting that $p(A|Z,C,\varphi) = 1$ for any pair $(Z,C) \in \Omega$, the following identity can be easily derived:

$$\frac{1}{p(A)} = \frac{p(Z,C,\varphi|A)}{p(Z,C,\varphi)}, \qquad \forall (Z,C) \in \Omega. \tag{25}$$

Using this identity in equation (24) we obtain:

$$\begin{aligned}\frac{1}{p(A)} &= \frac{1}{|\Omega|}\sum_{(Z,C)\in\Omega}\int \frac{p(\varphi)}{p(Z,C,\varphi)}p(Z,C,\varphi|A)d\varphi \\ &= \frac{1}{|\Omega|}\sum_{(Z,C)\in\Omega}\int \frac{1}{p(Z,C|\varphi)}p(Z,C,\varphi|A)d\varphi \\ &= \frac{1}{|\Omega|}\mathbb{E}\left[\frac{1}{p(Z,C|\varphi)}\Big|A\right].\end{aligned} \tag{26}$$

Consequently, using equation (26) we can estimate $p(A)$ as follows:

$$\widehat{p}(A) = \frac{|\Omega|}{\frac{1}{m}\sum_{l=1}^{m}\frac{1}{p(Z^{(l)},C^{(l)}|\varphi^{(l)})}}, \tag{27}$$

where each triplet $\left(Z^{(l)}, C^{(l)}, \varphi^{(l)}\right)$ is drawn from the posterior distribution $p(Z,C,\varphi|A)$. Therefore, this estimator corresponds to the harmonic mean of the likelihood of the augmented choices and coupons amplified by $|\Omega|$, where the values for $\left(Z^{(l)}, C^{(l)}, \varphi^{(l)}\right)$ can be obtained from the MCMC output.

Finally, we note that if two models $\mathcal{M}_1$ and $\mathcal{M}_2$ share the same set of feasible combinations of choices and coupons (i.e., $\Omega_{\mathcal{M}_1} = \Omega_{\mathcal{M}_2} = \Omega$), then for the purposes of model selection, it is not necessary to compute $|\Omega|$, which is constant for these two models and, thus, it does not affect the marginal-likelihood ratio $p(A|\mathcal{M}_1)/p(A|\mathcal{M}_2)$.

# 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] 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.

[4] —, 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.

[5] Bell, David R. and Jeongwen Chiang (2001), "Coupons and Consumer Budgets: A Within and Trans-Category Model of Shopping Behavior," Working Paper, University of Pennsylvania.

[6] Ben Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

[7] Berry, Steven, James Levinsohn and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4): 841-890.

[8] 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.

[9] Chen, Yuxin and Sha Yang (2004), "Estimating Disaggregate Model Using Aggregate Data via Augmentation of Individual Choice," Working Paper, New York University.

[10] Chiang, Jeongwen (1995), "Competing Coupon Promotions and Category Sales," *Marketing Science*, 10 (Fall): 297-315.

[11] 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.

[12] 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.

[13] 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.

[14] Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (1995), *Bayesian Data Analysis*. London: Chapman & Hall/CRC.

[15] —, Xiao-Ling Meng and Hal S. Stern (1996), “Posterior predictive assessment of model fitness via realized discrepancies,” *Statistica Sinica*, 6, 733–807.

[16] Kass, Robert E. and Adrian E. Raftery (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 90 (430) : 773-795.

[17] 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.

[18] Lenk, Peter (1992), “Hierarchical Bayes Forecasts of Multinomial Dirichlet Data Applied to Coupon Redemtions,” *Journal of Forecasting*, 11 : 603-619.

[19] 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.

[20] Little, Roderick J.A. and Donald B. Rubin (1987), *Statistical Analysis with Missing Data*, New York: John Wiley.

[21] 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.

[22] McCulloch, Robert E. and Peter E. Rossi (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64: 207-240.

[23] Musalem Andrés, Eric T. Bradlow and Jagmohan S. Raju (2004), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," Working Paper, University of Pennsylvania.

[24] Neslin, Scott A. (1990), "A market response model for coupon promotions," *Marketing Science*, 9 (Spring): 125-145.

[25] 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.

[26] Newton, Michael A. and Adrian E. Raftery (1994), "Approximating Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society (B)*, 56: 3-48.

[27] 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.

[28] Reibstein, David J. and Phillis A. Traver (1982), "Factors Affecting Coupon Redemption Rates," *Journal of Marketing*, 46 (Fall): 102-113.

[29] 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.

[30] 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.

[31] Ward, Ronald W. and James E. Davis (1978), "A Pooled Cross-Section Time Series Model of Coupon Promotions," *American Journal of Agricultural Economics*, 60 (August): 393-401.

[32] Yang, Sha, Yuxin Chen and Greg M. Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics* 1 (3): 251-275.