

# Purchase-Frequency Bias in Random-Coefficients Brand-Choice Models

Anand V. BODAPATI

Anderson Graduate School of Management, University of California, Los Angeles, CA 90095 (*bodapati@ucla.edu*)

Sachin GUPTA

Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853 (*sg248@cornell.edu*)

Conventional random-coefficients models of conditional brand choice using panel data ignore the dependence of the random-coefficients distribution on the purchase frequencies. We show that this leads to biased estimates and propose a conditional likelihood approach to obtain unbiased estimates. Unlike alternative approaches that require observation of “no-purchase” occasions, our proposed method relies only on purchase data. Furthermore, our approach does not require that the researcher specify the distribution of purchase frequencies. As a result, estimates of the brand-choice model are unaffected by misspecification of the model of purchase frequencies. We demonstrate the performance of the proposed approach in simulated data and in scanner data. We find that results differ substantively from the conventional latent-class model in terms of segment membership probabilities, segment characteristics, and price elasticities.

**KEY WORDS:** Choice model; Estimation method; Latent class model; Multinomial logit; Panel data; Random coefficient; Unobserved heterogeneity.

## 1. INTRODUCTION

A large body of marketing literature is based on statistical modeling of consumer brand-choice behavior using household panel data. In panel data we typically observe varying numbers of category purchases made by households during any observed time window. These differences in numbers of purchases reflect differences in category consumption (e.g., heavy buyers vs. light buyers) as well as the data-capture mechanism. For instance, panel data gathered via the frequent-shopper card of a particular retailer are likely to show more occurrences of purchases of consumers who are loyal to that retailer's stores. In this article we examine the implications of these differences in observed purchase frequencies of panelists on the modeling of brand-choice behavior.

Typically, researchers who only are interested in brand-choice behavior develop brand-choice models that account for heterogeneity in choice parameters between panelists, but ignore the process that generates differences in purchase frequencies. A partial list of such models includes those of Kamakura and Russell (1989), Chintagunta, Jain, and Vilcassim (1991), Gonul and Srinivasan (1993), and Fader and Hardie (1996). We show in this article that if the purchase frequency of a household is not independent of the parameters describing its choice probabilities, then current estimation approaches lead to biased estimates of the parameters of the choice model. We term this bias “purchase frequency bias.” In numerical simulations we show that the magnitude of this bias may be quite severe. Our result applies to both the primary models of unobserved heterogeneity prevalent in the literature—the Gaussian continuous random-effects model, and the finite-mixture or latent-class model.

Should we expect the purchase frequencies and brand-choice parameters to be dependent across panelists? Findings in the literature suggest this is indeed the case. For the canned tuna fish category, Kim and Rossi (1994) reported the following: “Our most striking finding is that consumers with high purchase frequency or high purchase volume are much more price sensitive and have more sharply defined preferences for national

brands than consumers with low frequency or low volume of purchase.” Dillon and Gupta (1996) found a positive relationship between purchase frequency and brand-choice price sensitivity in the paper towel category. In other purchase contexts, the relationship between purchase frequency and price sensitivity may be negative. For example, a segmentation study of customers of Mobil Oil (Forsyth, Gupta, Halder, Kaul, and Kettle 1999) found that price shoppers spent an average of \$700 annually, whereas price-insensitive, heavier users spent as much as \$1,200. Similarly, Chib, Seetharaman, and Strijnev (2004) found that the parameters explaining category purchases are correlated with the parameters explaining brand-choice decisions.

To illustrate the purchase frequency bias, we consider the following simple brand-choice structure. On each purchase occasion, consumers choose between two brands, A and B. Consumers are heterogeneous in their preferences for the two brands and belong to one of two homogeneous segments. Each consumer in segment 1 has a .30 probability of choosing brand A on each purchase occasion, whereas each consumer in segment 2 has a .70 probability of choosing brand A. For simplification, we omit marketing variables from this example. The two segments have an equal number of consumers. However, consumers in segment 1, who prefer brand B, tend to be light buyers and make four purchases in the category during the period studied in the panel data, whereas consumers in segment 2, who prefer brand A, are heavy buyers and make eight purchases.

The question of interest is as follows: If a latent-class model of brand choice (e.g., Kamakura and Russell 1989) is estimated on these data, would we recover the true brand preferences of each segment and the relative sizes of the segments? Surprisingly, the answer turns out to be “no.” Based on simulated data

for sample sizes as large as 10,000 panelists, the maximum likelihood estimates (MLEs) of the latent-class logit model and the true parameter values in this example are as follows:

	True value	MLE (standard errors) for the conventional latent-class logit model
Brand A choice probability in segment 1	.30	.21 (.008)
Brand A choice probability in segment 2	.70	.66 (.003)
Proportion of households in segment 1	.50	.29 (.010)

These results suggest that the estimates are biased and would mislead us about the segmentation structure in this market. In this article we explain why the MLEs in this example are biased, and in general the MLEs of random-coefficients brand-choice models are biased in many applications to real-world scanner panel data. Essentially, the explanation rests on the fact that the likelihood function for the model, as conventionally written, incorrectly assumes that the prior distribution of the brand-choice parameters is the same for all consumers, regardless of the length of their observed purchase strings. This assumption is clearly violated in the example presented previously, because consumers who purchase more frequently are more likely to have a preference for brand A. In this article we show that when the assumption is not met in the data, maximization of the conventional likelihood function produces inconsistent estimates.

In addition to identifying the purchase-frequency bias, a second contribution of this article is to propose an approach to eliminate the bias in conditional choice models. In essence,

to solve the problem, one needs to model the dependence between choices and purchase frequencies. The literature describes two approaches to modeling this dependence. One of these approaches is to jointly model brand-choice and category-purchase incidence (Chib et al. 2004; Chintagunta 1993; Chiang 1991) or purchase timing (Chintagunta and Prasad 1998), and the other is to jointly model brand-choice and purchase frequencies (Dillon and Gupta 1996). We propose a “conditional likelihood” approach to modeling the dependence. Figure 1 depicts the relationship between our approach and the literature. As we explain briefly next and in more detail later, the conditional likelihood approach that we develop is attractive relative to these other approaches in that it requires *only* conditional choice data and demands weaker assumptions on model structure and function forms.

“Conditional choice data” are data for which panelists’ behavior is observed only when they make a purchase in the category of interest. To jointly model brand choice and category purchase incidence, panelists need to be observed also on non-purchase occasions, that is, occasions when they potentially could have made a purchase in the category but did not. The data must also contain marketing variable values of all alternatives in the choice set on both purchase and nonpurchase occasions. Although a few academic studies have used no-purchase data, most clients of scanner data providers do not have access to such data.

To model purchase frequencies jointly with brand choice, conditional choice data are adequate as long as the purchase frequency parameters do not depend on the marketing mix variables. Dillon and Gupta (1996) offered such a model. Their approach is especially relevant when the researcher is interested in *both* brand-choice and purchase-frequency behaviors. However, the joint approach requires that the researcher choose specifications for both the brand-choice model and the distribution of

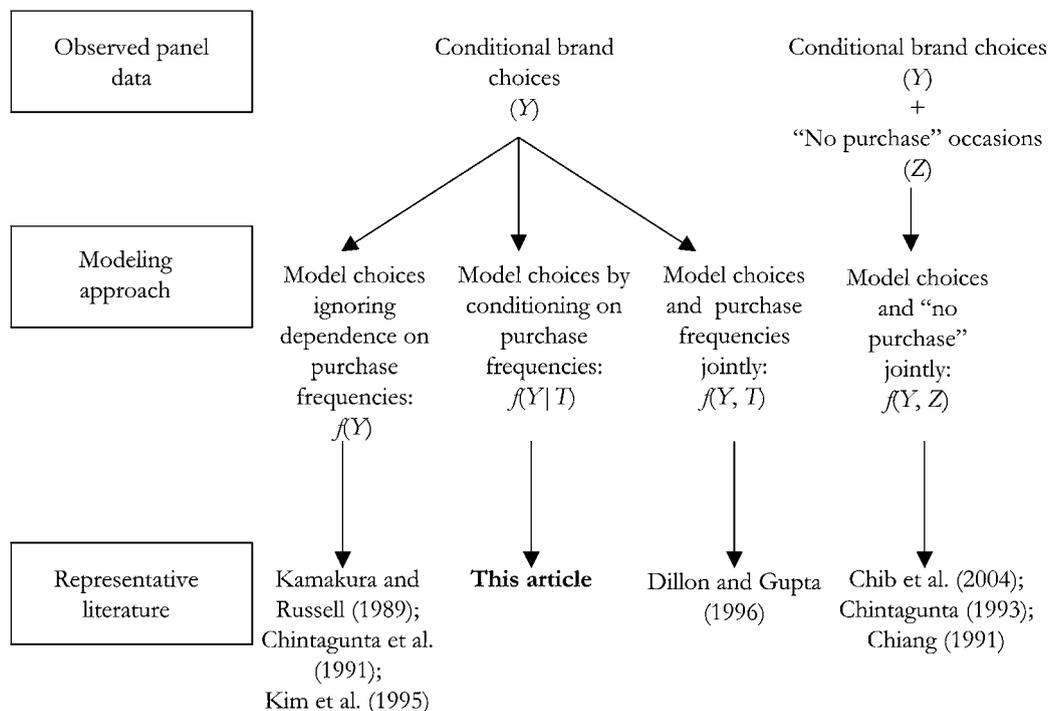


Figure 1. Relationship to the Literature.

purchase frequencies. Our objective, in contrast, is to obtain unbiased estimates of brand-choice parameters by controlling for their dependence on purchase frequencies. We do not aim to estimate the parameters for the purchase-frequency distributions. Our semiparametric conditional likelihood approach allows the researcher to remain relatively agnostic about the distribution of purchase frequencies.

To summarize, in this article we make two contributions to the brand-choice literature. First, we identify a purchase-frequency bias that is inherent in the estimation of random-coefficients model with household panel data. A large number of studies are susceptible to this bias. Second, we propose a conditional likelihood approach to correct the bias that offers advantages relative to extant methods. We demonstrate the significance of our findings by applying the latent-class brand-choice model and our proposed approach to household panel data on peanut butter purchases. Our results show that parameter estimates, price elasticities, and segment classifications are substantively and significantly altered when dependence on purchase frequencies is incorporated.

The rest of this article is organized as follows. In Section 2 we discuss the modeling framework, as well as the implications of the dependence of brand choice and purchase frequencies for maximum likelihood estimation of brand-choice parameters. In Section 3 we develop the conditional likelihood approach to obtain consistent estimates, and also present the joint likelihood approach for comparison. In Section 4 we use numerical experiments to demonstrate the inconsistency of extant estimation methods and the performance of our proposed approach. We also contrast performance of the conditional and joint approaches. In Section 5 we use panel data on peanut butter purchases to demonstrate the difference in results between the extant approach and the proposed approach. In Section 6 we summarize and conclude.

## 2. RANDOM-COEFFICIENTS MODELS AND INCONSISTENCY OF CURRENT ESTIMATION APPROACHES

In panel data we observe for household  $h$  the value of  $Y_{ht}$ , an index of the brand chosen on purchase occasion  $t$ ,  $t = 1, 2, \dots, T_h$ . We also observe a vector of values of marketing variables for each brand in the choice set on this purchase occasion,  $\mathbf{X}_t$ . Define  $\Pr(Y_{ht} = y_{ht} | \mathbf{X}_t, \boldsymbol{\beta})$  as the probability that household  $h$  makes a brand choice  $y_{ht}$  on occasion  $t$ , where  $\boldsymbol{\beta}$  is a vector of parameters that includes brand-specific constants and responsiveness to marketing mix variables. Using a random-utility framework and making certain assumptions on the distribution of the nondeterministic component of utility,  $\Pr(Y_{ht} = y_{ht} | \mathbf{X}_t, \boldsymbol{\beta})$  may be derived as a multinomial logit or a multinomial probit model, for instance. Define  $\mathbf{Y}_h$  as the vector whose components are  $Y_{h1}, Y_{h2}, \dots, Y_{hT_h}$ . Define  $p(\mathbf{Y}_h = \mathbf{y}_h | \mathbf{X}, \boldsymbol{\beta})$  as the corresponding likelihood of a given string of  $T_h$  brand choices of household  $h$ . This probability is the product of the choice probabilities on the  $T_h$  purchase occasions, under the assumption that the brand choices on separate occasions are conditionally independent, given the brand-choice parameters and the marketing-mix variables.

Next we introduce unobserved heterogeneity via a random-coefficients (RC) model. Define  $p(\boldsymbol{\beta} | \Gamma)$  as the distribution of household parameters,  $\boldsymbol{\beta}$ , in the population, where  $\Gamma$  denotes the parameters of the heterogeneity distribution.  $p(\boldsymbol{\beta} | \Gamma)$  may be assumed to be a continuous distribution such as the multivariate normal or a finite mixture of point densities, with the latter assumption resulting in the well-known latent-class model (Kamakura and Russell 1989). MLEs of  $\Gamma$  are obtained by maximizing the likelihood function obtained by taking the product of household-level likelihoods of the following form (suppressing the dependence on  $\mathbf{X}$  for simplicity)

$$L(\mathbf{Y}_h) = \int p(\mathbf{Y}_h | \boldsymbol{\beta}) p(\boldsymbol{\beta} | \Gamma) d\boldsymbol{\beta}. \quad (1)$$

In practice, the likelihood function in (1) often cannot be expressed in closed form for most continuous representations of  $p(\boldsymbol{\beta} | \Gamma)$ . Consequently, numerical integration methods or simulation methods (Brownstone and Train 1999) are often used to evaluate and maximize the likelihood function.

### 2.1 Inconsistency of the Above Maximum Likelihood Estimates

The likelihood function in (1) assumes that the distribution of  $\boldsymbol{\beta}$  is statistically independent of the length of households' purchase strings. To see exactly where this assumption is hidden, we need to consider the full argument of the likelihood function. We must first make explicit the fact that in writing down the likelihood of  $\mathbf{Y}_h$ , we take as exogenous and fixed the number of choice occasions,  $T_h$ . Therefore, the likelihood of interest, strictly speaking, is  $L(\mathbf{Y}_h | T_h)$ . This can be written as

$$L(\mathbf{Y}_h | T_h) = \int p(\mathbf{Y}_h | T_h, \boldsymbol{\beta}) p(\boldsymbol{\beta} | T_h, \Gamma) d\boldsymbol{\beta} \quad (2)$$

If the distribution of  $\boldsymbol{\beta}$  does not depend on  $T_h$ , then  $p(\boldsymbol{\beta} | T_h, \Gamma)$  reduces to  $p(\boldsymbol{\beta} | \Gamma)$ , and the likelihood in (2) reduces to that in (1). However, in general we expect the distribution of  $\boldsymbol{\beta}$  to depend on the number of purchase occasions. Therefore, the reduction of (2) to (1) ceases to be valid. This results in the likelihood in (1) becoming incorrect, and maximization of this likelihood function leads to inconsistent estimates of the true parameters. We call this bias the *purchase-frequency bias*.

The purchase-frequency bias occurs not only with maximum likelihood estimation, but also with Bayesian methods either entailing direct integration over the prior density or entailing Markov chain Monte Carlo methods. This is because as long as the Bayesian hyperprior density has non-0 value for all acceptable parameter values, the mean of the posterior density and the MLE both converge to the same point. Therefore, if the MLE is inconsistent, then so is the corresponding Bayesian estimate.

## 3. CORRECTING THE PURCHASE-FREQUENCY BIAS

We can identify two distinct approaches to correct the purchase-frequency bias when one has access only to conditional choice data:

1. The joint likelihood approach, which attempts to model  $L(\mathbf{Y}_h, T_h)$ , the joint density of  $\mathbf{Y}_h$  and  $T_h$

- The conditional likelihood approach, which directly models the correct conditional likelihood [which we see in (2)], possibly motivated by some underlying conceptualization of the joint density of  $\mathbf{Y}_h$  and  $T_h$ .

As we argue later in this section, the conditional likelihood approach may be preferred on the grounds that it makes weaker assumptions about the densities involved, presents a smaller number of parameters to be estimated, and lends itself more easily to a semiparametric formulation that is relatively agnostic about the form of certain density functions.

Here we focus primarily on the case where the distribution of  $p(\boldsymbol{\beta}|\Gamma)$  is a finite mixture of point densities. For this case, we discuss both the conditional likelihood and the joint likelihood approaches. Analogous development of the conditional and joint likelihood approaches when the distribution of  $p(\boldsymbol{\beta}|\Gamma)$  is continuous remains an area for future research. In Section 3.4 we discuss a method to modify the data to overcome the bias in the continuous-distribution case.

### 3.1 A Conditional Likelihood Approach for the Latent-Class Model

In the latent-class model (Kamakura and Russell 1989),  $p(\boldsymbol{\beta}|\Gamma)$  is a finite mixture of point densities. The model is motivated by assuming that there exist  $S$  segments of households,  $s = 1, 2, \dots, S$ , in the market under investigation. Each segment is characterized by its parameter vector,  $\boldsymbol{\beta}_s$ . For this model, the likelihood function in (2) becomes

$$L(\mathbf{Y}_h|T_h) = \sum_s p(\mathbf{Y}_h|T_h, \boldsymbol{\beta}_s)p(h \in s|T_h), \tag{3}$$

where  $p(h \in s|T_h)$  is the probability that household  $h$  belongs to segment  $s$ . If  $p(h \in s|T_h) = \pi_s$ , a constant not dependent on  $T_h$ , then we get the conventional likelihood function for the latent-class model,

$$L(\mathbf{Y}_h|T_h) = \sum_s p(\mathbf{Y}_h|T_h, \boldsymbol{\beta}_s)\pi_s. \tag{4}$$

To determine  $p(h \in s|T_h)$  when the independence assumption does not hold, define the population distribution of the number of purchase occasions across all households in segment  $s$  to be  $p(T_h|h \in s)$ . A simple application of Bayes's rule here tells us that given a certain  $T_h$ , the probability of the household belonging to segment  $s$  is

$$P(h \in s|T_h) = \frac{P(T_h|h \in s) \cdot \pi_s}{\sum_{s'} P(T_h|h \in s') \cdot \pi_{s'}}. \tag{5}$$

Substituting this expression for  $P(h \in s|T_h)$  into (3), we obtain the likelihood for the choices of household  $h$  observed on the  $T_h$  occasions as

$$L(\mathbf{Y}_h|T_h) = \sum_s L(\mathbf{Y}_h|\boldsymbol{\beta}_s, T_h) \frac{P(T_h|h \in s) \cdot \pi_s}{\sum_{s'} P(T_h|h \in s') \cdot \pi_{s'}}. \tag{6}$$

Equation (6) represents a generalized likelihood for the observed choices of a household in the situation where the number of purchase occasions is treated as exogenous. Therefore, using (6) as the basis for maximum likelihood estimation will yield consistent estimates for the segment membership probabilities  $\pi_s$  and the segment characteristics  $\boldsymbol{\beta}_s$ . To complete the

specification of the likelihood in (6), we need to choose a form for the term  $P(T_h|h \in s)$  that appears in it. We first discuss, in Sections 3.1.1 and 3.2, the form of the conditional likelihood obtained when one assumes that the analyst has knowledge about the functional form for the density  $P(T_h|h \in s)$ . In Section 3.3 we develop a semiparametric method that assumes weaker knowledge about this density on the part of the analyst.

*3.1.1 Distribution of Number of Purchase Occasions Within Segments.* The analyst must choose a density function form  $f(T_h; \boldsymbol{\tau}_s)$  that can suitably characterize the distribution of the number of purchase occasions within a segment. The  $\boldsymbol{\tau}_s$  represents the parameters of the density function. For instance, in certain situations we may consider the negative binomial distribution (NBD) to be a suitable characterization of the number of purchase occasions within a segment, whereas in other cases we may choose the geometric or Poisson distribution as appropriate. For ease of exposition of the expressions that follow, we give the density form of these three distributions expressed in a manner that brings to attention the form of the log-density:

$$\text{NBD: } f(T_h; \boldsymbol{\tau}_s) = \exp(\tau_{0s} + \tau_{1s}T_h - \ln(T_h!)) + \ln((\tau_{2s} + T_h - 1)!). \tag{7}$$

$$\text{Geometric: } f(T_h; \boldsymbol{\tau}_s) = \exp(\tau_{0s} + \tau_{1s}T_h). \tag{8}$$

$$\text{Poisson: } f(T_h; \boldsymbol{\tau}_s) = \exp(\tau_{0s} + \tau_{1s}T_h + T_h!). \tag{9}$$

Let us write the density in general of the number of purchase occasions within a segment as

$$f(T_h; \boldsymbol{\tau}_s) = \exp(\tau_{0s} + m(T_h; \tau_{1s}, \dots, \tau_{ns})). \tag{10}$$

The  $\tau_{0s}$  in the expressions (7)–(10) represents a normalizing constant that is a strict function of the other  $\tau$ 's, so that each density's values sum to 1 over  $T_h = 0, 1, \dots$ . If the log-density is linear in the parameters  $\tau_{0s}, \tau_{1s}, \dots, \tau_{ns}$ , then the density form belongs to the exponential family of distributions (Mood, Graybill, and Boes 1974). The exponential family is a large family that accommodates a diverse set of distributions, including the normal, Poisson, and geometric, and takes the general form

$$f(T_h; \boldsymbol{\tau}_s) = \exp(n(\boldsymbol{\tau}_s) + \tau_{1s}m_1(T_h) + \tau_{2s}m_2(T_h) + \dots + \tau_{ns}m_n(T_h)), \tag{11}$$

where  $\boldsymbol{\tau}_s = (\tau_{1s}, \dots, \tau_{ns})$  are the parameters of the density and  $(m_1, \dots, m_n)$  are the smooth basis functions of  $T_h$  that essentially characterize the nature of a particular member of the exponential family. Here  $n(\boldsymbol{\tau}_s)$  is equivalent to  $\tau_{0s}$  in (10), because it is the normalizing constant depending strictly on the other parameters  $(\tau_{1s}, \dots, \tau_{ns})$ . As expression (7) shows, the NBD does not have a log-density that is linear in the parameters.

Linearity of the log-density as the exponential family affords has computational consequences that we mention later. Let us return to the general density form of (10). From (6), we get the fully parameterized likelihood function for household  $h$  as

$$\begin{aligned} L(\mathbf{Y}_h|T_h) &= \sum_s L(\mathbf{Y}_h|\boldsymbol{\beta}_s, T_h) \\ &\times \frac{\exp(\tau_{0s} + m(T_h; \tau_{1s}, \dots, \tau_{ns})) \cdot \pi_s}{\sum_{s'} \exp(\tau_{0s'} + m(T_h; \tau_{1s'}, \dots, \tau_{ns'})) \cdot \pi_{s'}} \end{aligned}$$

$$\begin{aligned}
&= \sum_s L(\mathbf{Y}_h | \boldsymbol{\beta}_s, T_h) \\
&\quad \times \frac{\exp(\tau_{0s} + \ln(\pi_s) + m(T_h; \tau_{1s}, \dots, \tau_{ns}))}{\sum_{s'} \exp(\tau_{0s'} + \ln(\pi_{s'}) + m(T_h; \tau_{1s'}, \dots, \tau_{ns'}))} \\
&= \sum_s L(\mathbf{Y}_h | \boldsymbol{\beta}_s, T_h) \\
&\quad \times \frac{\exp(\tau'_{0s} + m(T_h; \tau_{1s}, \dots, \tau_{ns}))}{\sum_{s'} \exp(\tau'_{0s'} + m(T_h; \tau_{1s'}, \dots, \tau_{ns'}))}. \quad (12)
\end{aligned}$$

We collapse  $(\tau_{0s} + \ln(\pi_s))$  into  $\tau'_{0s}$ , because the two terms are not separately identifiable without imposing certain restrictions on those terms. The likelihood for all of the observed choices in the panel is the product of the foregoing term over all households,

$$L(H) = \prod_h L(\mathbf{Y}_h | T_h).$$

Maximizing this expression over  $(\{\theta_{1s}, \theta_{2s}, \dots, \theta_{K-1s}, \boldsymbol{\beta}_s, \tau'_{0s}, \tau_{1s}, \dots, \tau_{ns}\}_{s=1}^S)$  yields consistent estimates of those quantities.

**3.1.2 Estimating the Segment Frequencies  $\pi_1, \pi_2, \dots, \pi_s$ .** We now briefly consider the issue of estimating the fraction of households in the population that belong to any given segment. Given the MLEs of  $(\{\tau'_{0s}, \tau_{1s}, \dots, \tau_{ns}\}_{s=1}^S)$ , we can obtain MLEs of  $P(h \in s | T_h)$  by plugging the appropriate values into the expression

$$P(h \in s | T_h) = \frac{\exp(\tau'_{0s} + m(T_h; \tau_{1s}, \tau_{2s}, \dots, \tau_{ns}))}{\sum_{s'} \exp(\tau'_{0s'} + m(T_h; \tau_{1s'}, \tau_{2s'}, \dots, \tau_{ns'}))}.$$

The foregoing expression can be seen to give the *household-level* prior probabilities on segment membership. A consequence of the self-consistency principle of maximum likelihood estimation is that the across-household average of the household-level prior probability for segment  $s$  is the same as the population-level prior probability for segment  $s$ , which is  $\pi_s$ . Therefore, we may estimate the segment frequency  $\pi_s$  by the empirical average of the foregoing expression evaluated at the  $T_h$  value for each household.

**3.1.3 Concomitant-Variable Model.** It is interesting to note here the similarity between the expression in (12) and the concomitant-variable latent-class logit model (Dayton and Macready 1988; Gupta and Chintagunta 1994). The concomitant-variable model was introduced as a way to include individual-specific characteristics such as income, age, and household size, to characterize the probability of segment membership in a latent-class model. The expression in (12) may be viewed analogously as a concomitant-variable model where the individual-specific characteristics are some functions of the number of purchase occasions  $T_h$ . For the exponential family of distributions, the translation is apparent; the individual-specific characteristics are simply the basis functions  $(m_1, \dots, m_n)$ . Given this similarity, we henceforth refer to the model in (12) as the *concomitant function model*, and the term

$$(\tau'_{0s} + m(T_h; \tau_{1s}, \dots, \tau_{ns})) \quad (13)$$

as the *concomitant function* (CF). For the exponential family, the CF is linear in the parameters. This results in a simple form for the gradient of the likelihood function with respect to the parameters  $(\tau_{1s}, \dots, \tau_{ns})$ , which simplifies maximization of the likelihood function when one uses a gradient-based optimization algorithm.

**3.1.4 Choice of Basis Functions in the Concomitant Function.** Comparing (7) and (10) we see that the basis functions in the log-density for the negative binomial distribution are  $T$ ,  $\ln(T!)$ , and  $\ln((\tau_{2s} + T - 1)!)$ . Similarly, comparing expressions (8) and (10), we see that the basis function for the geometric density is  $T$ . In general, how does the analyst choose the appropriate basis functions [and, consequently, the form of the density  $P(T_h | h \in s)$ ]? Clearly, if the analyst knows that the distribution is the NBD or some known member of the exponential family, then the analyst will choose the corresponding basis functions for use in the likelihood function (12). However, in most empirical applications one rarely knows the true density, and so this approach may not be workable. Therefore, a semi-parametric approach, whereby we choose a set of basis functions that is flexible enough to accommodate a wide range of densities, may be attractive. The choice of basis functions is a common problem in approximation theory, and, depending on the application, one may use monomials, Legendre polynomials, Fourier sinusoids, and such. We recommend simple monomials for their ease of interpretability and flexibility, and so we choose

$$\begin{aligned}
m_1(T_h) &= T_h, \\
m_2(T_h) &= T_h^2, \\
m_3(T_h) &= T_h^3, \dots
\end{aligned} \quad (14)$$

To examine the flexibility of the monomial basis functions, we approximated the binomial and NBDs with monomial basis functions of second and third degree for 1,000 combinations of means and variances. The 1,000 samples were rank-ordered based on goodness of fit ( $R^2$ ). For the quadratic function, goodness of fit at the 50th percentile exceeds .99 for the binomial and .95 for the NBD. For the cubic function, goodness of fit at the 50th percentile exceeds .99 for the NBD as well. The goodness of fit in absolute terms is very high even for cases where the fit is poor relative to other cases, as we can see from the corresponding numbers at the 10th percentile. For the quadratic function, goodness of fit at the 10th percentile exceeds 98% for the binomial and 87% for the NBD. For the cubic function, goodness of fit at the 10th percentile exceeds 99% for the binomial and exceeds 96% for the NBD. In Section 4 we show via numerical simulations that monomial basis functions perform very well when applied to panel data.

## 3.2 A Joint Likelihood Approach for the Latent-Class Model

The joint model entails developing the joint probability of observing the brand-choice vector  $Y_{ht}$  and the purchase frequency  $T_h$ . We represent this probability by  $p(\mathbf{Y}_h = \mathbf{y}_h, T_h = t_h | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\tau})$ , where  $(\boldsymbol{\beta}, \boldsymbol{\tau})$  is the vector of parameters of the joint distribution. As before, let  $p(\boldsymbol{\beta}, \boldsymbol{\tau} | \Gamma)$  represent the unobserved heterogeneity distribution of parameters in the population of households. Integrating out the heterogeneity distribution, the joint likelihood function for household  $h$  may be written as

$$L(\mathbf{Y}_h, T_h) = \int p(\mathbf{Y}_h, T_h | \boldsymbol{\beta}, \boldsymbol{\tau}) p(\boldsymbol{\beta}, \boldsymbol{\tau} | \Gamma) d\boldsymbol{\beta} d\boldsymbol{\tau}. \quad (15)$$

Focusing on the case where the heterogeneity distribution is a finite mixture of point densities, and assuming that brand

choice and purchase frequencies are independent conditional on a household's membership in a latent class, we obtain the following likelihood function:

$$L(\mathbf{Y}_h, T_h) = \sum_s p(\mathbf{Y}_h | \boldsymbol{\beta}_s) p(T_h | \boldsymbol{\tau}_s) p(h \in s). \quad (16)$$

This likelihood function corresponds with the development of Dillon and Gupta (1996), with  $p(T_h | \boldsymbol{\tau}_s)$  chosen to be a truncated-at-1 Poisson density.

It can be shown that maximizing the foregoing joint likelihood will in general yield more efficient estimates than maximizing the conditional likelihood. This is because the joint likelihood approach in effect draws inferences about the parameters not only via the conditional density of  $\mathbf{y}$  given  $T_h$  (which is what the conditional approach does), but also via the marginal density of  $T_h$ . However, the usual motif in estimation theory holds, that higher efficiency typically comes at the expense of more assumptions on structure and density form. We see this here as well; the joint likelihood approach requires us to posit a certain form for the joint density  $p(\mathbf{Y}_h, T_h)$ , whereas the conditional approach requires us to posit a form for only the conditional density  $p(\mathbf{Y}_h | T_h)$ . Whereas a choice of functional form for the joint density uniquely determines the functional form for the conditional density, the converse is not true. In this sense, one makes stronger assumptions about model form by positing a joint density than by positing a conditional density. In fact, one can show that for any (segment-level) exponential family density function  $p(T_h | \boldsymbol{\tau}_s)$  with a sufficient-statistics vector of dimensionality  $n$  as in (11), the overall number of identifiable parameters in the conditional likelihood function (12) is  $n$  less than the number of identifiable parameters for the joint likelihood function (14). This is why the conditional approach recovers only the parameters related to segment size and brand choice, and not the parameters related to the purchase-frequency distribution. In this sense, too, the conditional approach makes less strong demands and requires that the analyst assume less about model structure.

As we stated earlier in Section 3.2, one rarely knows the true density functions involved. We made the remark in reference to the conditional density, but it also holds true—perhaps even more strongly—for the joint density. We attempted to address this issue for the conditional density by proposing a semiparametric approach in which the basis functions are chosen from a very flexible set like the monomials. It is natural to ask, whether a similar semiparametric approach can be developed for the joint density as well. The answer is “yes,” but its implementation is considerably more awkward and numerically more difficult. Consequently, this kind of semiparametric approach is less workable in the case of the joint density. To see where the trouble lies, consider the exponential family density form corresponding to the monomial basis functions

$$p(T_h | h \in s) = \exp(\tau_{0s} + \tau_{1s} T_h + \tau_{2s} T_h^2 + \tau_{3s} T_h^3 + \dots + \tau_{ns} T_h^n).$$

As we said earlier, the  $\tau_{0s}$  is just a normalizing constant. Therefore, it is not a separately identifiable parameter, but rather a function of the other parameters,

$$\tau_{0s} = -\log \sum_{T_h=1}^{\infty} \exp(\tau_{1s} T_h + \tau_{2s} T_h^2 + \tau_{3s} T_h^3 + \dots + \tau_{ns} T_h^n).$$

Closed-form simplifications of this expression are not possible even with moderately small-degree monomial basis functions; instead, one will need to resort to approximations that will typically be very nonlinear in the parameters. This expression makes the overall joint likelihood function [of the form in (16)] a very nonlinear function of the monomial parameters serving as the coefficients of the monomial terms. The practical consequence of this is that estimation procedures are often poorly behaved and ill-conditioned. The normalizing constants do not present a problem in the conditional likelihood approach, because in that situation they dissolve into the other parameters and the  $n$  parameters that are not estimated in this approach. For these reasons, the semiparametric monomial approach is less workable with the joint likelihood than with the conditional likelihood.

### 3.3 A Data-Based Solution Approach

In the continuous heterogeneity case, a model-based solution is not available yet and must await future research. Meanwhile, we propose a data-based “fix” to the purchase frequency bias. Because a necessary condition for the bias to arise is an unequal number of purchase observations across panelists, we propose randomly sampling an equal number of purchases from each panelist's available data. In the next section we examine the performance of this approach in numerical experiments.

## 4. NUMERICAL EXPERIMENTS

We conduct numerical experiments in which we generate data by systematically violating the independence assumption underlying the conventional likelihood function for the random-coefficients model. We use these simulated data to illustrate that parameter estimates obtained from current approaches to estimation of the random-coefficients model are biased. Using the same simulated data, we further show that our proposed conditional likelihood approach produces unbiased estimates. Finally, we show that although the joint likelihood approach produces unbiased results when the purchase-frequency density is specified correctly, it fails when the density is misspecified. (Results of additional simulation experiments confirm the robustness of the results reported in this article. These results may be obtained from the authors on request.)

### 4.1 Experimental Design

We simulate a market with 1,000 household panelists whose purchases in a category are recorded over a period of 100 weeks. On each purchase occasion, the household chooses to buy one of two brands, A and B, in the category, based on maximization of utility. The brands' utilities are iid extreme value-distributed; the utility is influenced by the price of the brand. Weekly prices of the two brands are drawn such that the price difference (price A – price B) is uniformly distributed on the interval  $(-1, 1)$ .

In the design of the simulation experiment we consider two factors, each of which is considered at two levels. The resulting four cells are labeled  $i, j$ , where  $i = 1, 2$  and  $j = 1, 2$ . The first factor is whether the extent of heterogeneity in the choice

process is low ( $i = 1$ ) or high ( $i = 2$ ). The second factor is whether the distribution of purchase frequencies is independent of the distribution of choice parameters ( $j = 1$ ) or not ( $j = 2$ ). This design is considered separately for discrete and continuous heterogeneity distributions.

**4.1.1 Discrete Heterogeneity Distribution.** Consistent with the latent-class conditional-choice model, households belong to one of two segments. Here 30%, or 300, households belong to segment 1, and the remaining 700 households belong to segment 2. Conditional on belonging to a segment, the household's brand specific constant for brand A (setting the brand-specific constant for brand B to 0) and price responsiveness are fixed. Thus there are four parameters in the choice process.

Within each choice segment, the number of category purchases  $T_h$  made by a panelist during this period is a draw from a truncated-at-1 NBD. Truncation serves the purpose of allowing only strictly positive integer values for the number of category purchases. One way to characterize the NBD is through two parameters ( $\tau_1, \tau_2$ ) that represent the mean and variance. After the number of purchases is determined for a household, the actual weeks when purchases occur are determined by sampling according to the uniform distribution from the 100 weeks, with replacement.

Table 1 shows the true values of parameters in the experiment. Higher heterogeneity in choice parameters (factor 1) is

obtained by increasing the spread between the two segments' parameters. A measure of higher heterogeneity is that the difference between the variance-covariance matrices ( $\mathbf{V}_{High} - \mathbf{V}_{Low}$ ) is positive definite. To obtain independence between the purchase frequencies and brand choice parameters (factor 2), we specify the distributions of purchase frequencies to be the same in segments 1 and 2. Conversely, to obtain dependence, we specify the two NBDs to have different means. Segment 2 is assumed to make 12.14 purchases on average, compared with 5 purchases made by segment 1. The values of segment-level means are chosen such that the average number of purchases in the market is 10 at both levels of factor 2.

**4.1.2 Continuous Heterogeneity Distribution.** We assume that  $(\alpha, \beta, \tau_1)$  are jointly multivariate normal. For comparability, the means and variances of  $(\alpha, \beta)$  for low and high heterogeneity are specified at the same values as implied by the parameter values in the discrete case described previously. We pick  $cov(\alpha, \beta)$  such that we induce mild correlation of .20 between  $\alpha$  and  $\beta$ . (In the discrete case with only two segments in the market, the correlation between  $\alpha$  and  $\beta$  is 1. Because this would be unrealistic for a continuous representation of heterogeneity, we chose the correlation to be .20.) To create dependence between purchase frequencies and the price sensitivity parameter, we chose values of  $cov(\beta, \tau_1)$  to induce correlation of .70. Note that because  $T_h$  is a random perturbation around  $\tau$ , the correlation between  $\beta$  and  $T_h$  is considerably

Table 1. Comparison of Performance of Alternative Models for Discrete Heterogeneity: Mean Parameter Estimates and (bias in number of standard errors) in 50 Sample Replications

Model	Chi-squared (p value)	$\alpha_1$	$\beta_1$	$\alpha_2$	$\beta_2$	$\pi_1$
Cell 1.1: Factor 1, low; factor 2, independent						
True values		1.5	-2.0	.5	-3.0	.300
Current method	4.76 (p = .45)	1.493 (.407)	-2.043 (1.724)	.477 (2.869)	-3.037 (2.134)	.321 (2.648)
Concomitant method	NA	1.491 (.555)	-2.042 (1.647)	.476 (2.921)	-3.037 (2.224)	.323 (2.989)
Joint with truncated Poisson	32.65 (p < .001)	1.365 (2.239)	-2.363 (2.716)	.599 (2.195)	-2.882 (2.111)	.314 (.152)
Cell 1.2: Factor 1, low; factor 2, dependent						
True values		1.5	-2.0	.5	-3.0	.300
Current method	38.98 (p < .001)	1.671 (3.078)	-1.753 (3.399)	.546 (4.939)	-2.960 (2.818)	.137 (10.724)
Concomitant method	NA	1.504 (.194)	-1.954 (1.797)	.511 (2.226)	-3.005 (.483)	.289 (2.336)
Joint with truncated Poisson	20.32 (p < .001)	1.516 (1.053)	-1.919 (3.211)	.565 (14.877)	-2.918 (10.410)	.326 (8.997)
Cell 2.1: Factor 1, high; factor 2, independent						
True values		1.5	-1.0	-.5	-3.0	.300
Current method	4.32 (p = .50)	1.494 (.657)	-.984 (1.079)	-.501 (.134)	-2.988 (1.167)	.300 (.245)
Concomitant method	NA	1.494 (.671)	-.985 (1.062)	-.501 (.134)	-2.988 (1.171)	.300 (.191)
Joint with truncated Poisson	37.80 (p < .001)	1.330 (2.140)	-1.152 (1.840)	-.334 (2.122)	-2.815 (2.333)	.574 (3.889)
Cell 2.2: Factor 1, high; factor 2, dependent						
True values		1.5	-1.0	-.5	-3.0	.300
Current method	29.33 (p < .001)	1.688 (10.913)	-.867 (5.348)	-.440 (15.007)	-2.910 (10.317)	.191 (67.676)
Concomitant method	NA	1.497 (.274)	-1.009 (.480)	-.498 (.357)	-2.990 (1.188)	.299 (.772)
Joint with truncated Poisson	100.92 (p < .001)	1.519 (1.596)	-.975 (1.295)	-.455 (8.748)	-2.924 (7.415)	.291 (7.240)

NOTE:  $(\tau_1, \tau_2)$  are mean and variance parameters of the NBD.  $\tau_2$  is equal to 13 in all cells. In cells 1.1 and 2.1,  $E(\tau_1) = 10$  in both segments. In cells 1.2 and 2.2,  $E(\tau_1) = 5$  in segment 1 and  $E(\tau_1) = 12.14$  in segment 2. NA, not applicable.

Table 2. Performance of the RC Model for Continuous Heterogeneity: Mean Parameter Estimates and (bias in number of standard errors) in 50 Sample Replications

Model	$E(\alpha)$	$E(\beta)$	$var(\alpha)$	$var(\beta)$	$cov(\alpha, \beta)$
Cell 1.1: Factor 1, low; factor 2, independent					
True values	.7	-2.7	.21	.21	.042
RC model on full data	.698 (.372)	-2.706 (.641)	.237 (1.231)	.199 (2.031)	.055 (1.696)
Cell 2.1: Factor 1, high; factor 2, independent					
True values	.1	-2.4	.84	.84	.168
RC model on full data	.098 (.287)	-2.403 (.261)	.874 (1.146)	.826 (1.763)	.162 (.464)
Cell 1.2: Factor 1, low; factor 2, dependent					
True values	.7	-2.7	.21	.21	.042
RC model on full data	.691 (1.775)	-2.677 (2.562)	.206 (.748)	.159 (3.192)	.045 (.431)
RC model on equal strings data	.699 (.140)	-2.739 (2.441)	.267 (1.381)	.236 (1.328)	.035 (.335)
Cell 2.2: Factor 1, high; factor 2, dependent					
True values	.1	-2.4	.84	.84	.168
RC model on full data	.096 (.739)	-2.341 (5.989)	.824 (1.526)	.648 (7.755)	.164 (.349)
RC model on equal strings data	.099 (.073)	-2.396 (.171)	.684 (1.903)	.897 (1.224)	.197 (.739)

NOTE:  $(\tau_1, \tau_2)$  are mean and variance parameters of the NBD.  $E(\tau_1) = 10$  and  $V(\tau_1) = 2.676$ .  $\tau_2$ , the variance parameter of the NBD, is held fixed at 13. In cells 1.2 and 2.2,  $cov(\beta, \tau_1)$  is chosen so that the implied correlation is .70.

smaller, approximately .14. The assumption of joint normality implies that the conditional distribution of the choice parameters  $(\alpha, \beta|\tau_1)$  continues to be normal. This is essential to ensure that the Gaussian random-coefficients model is correctly specified in the simulations. Parameter values are given in Table 2.

To obtain the sampling distribution of the estimator, we generate 50 sample replicates in each of the 4 cells of the simulation experiment for discrete and continuous heterogeneity cases separately. Thus, we generate a total of 400 datasets and use them for estimating each model of interest. Standard errors of the estimated mean value of the parameter are computed as  $s/\sqrt{n}$ , where  $s$  is the standard deviation of the estimate across 50 samples and  $n = 50$ . Bias is measured as the distance of the mean parameter estimate from the true parameter value in terms of number of standard errors. Estimated segment parameters are matched up with true parameters by visual inspection. To test whether a particular model yields unbiased parameter estimates, we use the classical likelihood ratio test discussed by, for instance, Bickel and Docksum (1977). Finally, we also compare results from predictions of market shares in hold-out samples, based on alternative models.

#### 4.2 Results From Current Estimation Methods

We estimated the classical latent-class model using maximum likelihood on all four cells in the discrete heterogeneity case. Results are given in Table 1 in the rows titled 'Current method.' As expected, the likelihood ratio test (chi-squared statistic) indicates that parameter estimates are unbiased when the distribution of purchase frequencies is the same across choice segments (cells 1.1 and 2.1). Moreover, the bias in each of the estimated parameters is small. The largest bias is just under three standard errors. This result confirms our belief that the current estimation method yields unbiased estimates under independence.

When the choice process is dependent on the purchase-frequency distribution (cells 1.2 and 2.2), the chi-squared test

rejects ( $p < .01$ ) the null hypothesis of unbiasedness. Moreover, bias in individual parameter estimates is large. The bias in the estimate of the segment size ( $\pi$ ) is particularly pronounced, with the bias reaching 67 standard errors in cell 2.2. The absolute bias in  $\pi$  is as large as 54% in cell 1.2. Importantly, estimates of  $\alpha$  and  $\beta$  are biased as well. For example, in segment 1 the bias in the price parameter is about 12% in each of the two cells, 1.2 and 2.2.

In the continuous heterogeneity case, we estimate the random-coefficients model under the assumption of normality of  $p(\beta|\Gamma)$  using the method of simulated maximum likelihood (Brownstone and Train 1999). Results are given in Table 2. As in the discrete case, under independence (cells 1.1 and 2.1) the bias in all parameters is small, with the largest bias just over two standard errors. Under dependence (cells 1.2 and 2.2), however, estimates of  $E(\beta)$  and  $var(\beta)$  are biased. The bias is quite severe when choice heterogeneity is high (cell 2.2).

#### 4.3 Results From Proposed Solution Approaches for Discrete Heterogeneity

Table 1 gives results from application of the concomitant model to data in the discrete heterogeneity cases. Our concomitant model is quadratic in the purchase frequency. The likelihood ratio test is not applicable here, because the estimated model is not nested within the true model. As expected, in cells 1.1 and 2.1 the performance of the concomitant model is extremely similar to the performance of the classical latent-class model. However, in cells 1.2 and 2.2 the concomitant model shows a dramatic improvement relative to the classical method. In these two cells we find that the bias is smaller than 2 standard errors in 8 of the 10 parameters and smaller than 2.5 for the remaining 2 parameters. Note that Table 1 does not report the coefficients of the concomitant functions; we report only those parameters that have equivalents across all of the models being estimated.

**4.3.1 Results From the Joint Likelihood Approach.** We apply the joint model of brand choice and purchase frequency to data in all four cells in the discrete heterogeneity case. When the estimation model is correctly specified, (i.e., the purchase frequency distribution is truncated-at-1 negative binomial), the model yields unbiased results in all four cells, as expected. The  $p$  values of the chi-squared test statistics are .45, .49, .47, and .46 in cells 1.1, 1.2, 2.1, and 2.2. For reasons of space, we do not provide detailed parameter estimates here.

Next, we consider the following question: Because a priori the analyst may not know the true distribution of purchase frequencies, what is the impact of misspecification on the estimated brand choice parameters? To study this, we specify the purchase frequency to be truncated-at-1 Poisson instead of negative binomial. Table 1 gives parameter estimates and chi-squared statistics from application of the misspecified joint model to cells 1.2 and 2.2. The likelihood ratio test resoundingly rejects the null hypothesis of unbiasedness. It is also clear that the bias in individual parameter estimates is quite large in terms of number of standard errors. In contrast, the quadratic concomitant model (also shown in Table 1) provides relatively unbiased results.

**4.3.2 Comparison of Models' Predictive Ability.** We examine the performance of all three models in predicting expected market shares in holdout samples in cells 1.2 and 2.2 in the discrete heterogeneity case. Recall that in these two cells, the distribution of purchase frequencies is *not* independent of the distribution of choice probabilities. In each experimental cell, we create 25 pairs of datasets randomly. In each pair, we use model estimates and (purchase rate-adjusted) segment sizes from one dataset to predict market shares in the other dataset. Table 3 presents two measures of predictive fit, mean absolute deviation (MAD) (weighted by true market share) and mean absolute percentage deviation (MAPD). Predictions of market shares require knowledge of both choice probabilities and purchase frequencies. This knowledge is assumed both for the current latent class model and for the proposed concomitant model, thereby putting the models on equal footing.

The results for out-of-sample predictions indicate the superiority of the concomitant model. The sources of poor performance of the joint model are twofold: (1) Purchase frequencies are predicted poorly because of model misspecification, and (2) choice probabilities are predicted poorly because parameter estimates are biased due to misspecification of the purchase frequency density. We speculate that the current method has smaller error, because although the choice probabilities are biased, the model uses the empirical distribution of purchase

frequencies for predicting market shares. Finally, the concomitant model yields unbiased choice probabilities and uses the empirical distribution of purchase frequencies. Consequently, it gives the best predictions of market shares.

#### 4.4 Results From the Data-Based Solution Approach

Table 2 also gives results from application of the data-based solution to data in cells 1.2 and 2.2 in the continuous heterogeneity case. To achieve (almost) equal purchase frequencies across panelists, in cell 2.2 we sample two observations (three in cell 1.2) from the available data for each panelist who made more than one purchase (more than two purchases in cell 1.2) in the original data. (About 4% of panelists buy exactly once, and about 1% buy exactly twice in the simulated data. Strictly speaking, the data-based solution would require sampling one observation from each panelist. However, when only one observation per panelist is included, the maximum likelihood procedure fails to converge for more than half of the datasets. With two purchases, we experience convergence problems in cell 1.2, but not in cell 2.2. Consequently, we sample two purchases in cell 2.2 and three purchases in cell 1.2.) Data for panelists who made only one purchase are included without changes in the new sample. As a consequence of this transformation, the sample size is reduced on average to less than 30% of the original sample size, because the mean purchase frequency was 10 in the original data. In cell 2.2, all five parameter estimates have  $t$ -values smaller than 2. In cell 1.2, the estimate of  $E(\beta)$  continues to have a  $t$ -statistic slightly larger than 2, but all other parameters are unbiased. Because in this cell our data-based solution reduces but does not eliminate the dependence, it is unsurprising that the purchase frequency bias persists to a small degree.

## 5. APPLICATION TO SCANNER PANEL DATA

We use household scanner panel data for the peanut butter category to illustrate the implications of recognizing the dependence between purchase frequencies and brand preferences and marketing mix responsiveness of panelists. These data were collected by the A. C. Nielsen Company in the Sioux Falls (SD) market. We select a subset of six items in the category to study. The selected items (Jif, 18 oz., 28 oz., and 40 oz.; Peter Pan, 18 oz.; and Skippy, 18 oz., 28 oz., and 40 oz.) have the largest market shares and together account for 60% of category volume in the market. All category purchases over a 102-week period of a randomly chosen sample of 829 households are used for model estimation, resulting in a sample size of 5,005 purchases.

Table 3. Results of Out-of-Sample Market Share Predictions for Discrete Heterogeneity Distribution

Cell	Model	MAPD brand 1 market share (% of true)	MAD brand 1 market share (% points)
1.2	Current method	1.13	.45
1.2	Concomitant method	.95	.38
1.2	Joint with truncated Poisson	5.30	2.10
2.2	Current method	4.89	1.23
2.2	Concomitant method	1.28	.32
2.2	Joint with truncated Poisson	9.01	2.28

Table 4. Descriptive Statistics of Peanut Butter Data

Item	Brand share %	Mean price (cents/ounce)	% of purchases on display	% of purchases on feature
Jif 18 oz.	17.3	9.76	0	2.5
Jif 28 oz.	9.9	9.97	0	2.5
Jif 40 oz.	2.8	9.88	0	0
Peter Pan 18 oz.	22.2	9.35	2.4	8.9
Skippy 18 oz.	31.2	9.36	3.8	13.2
Skippy 28 oz.	9.0	9.54	1.6	5.9
Skippy 40 oz.	7.7	9.65	.2	1.0

NOTE:  $n = 5,005$  purchase observations, 829 panelists.

The selected panelists made an average of 6 category purchases during the period under study (range, 1–56 purchases). Table 4 gives descriptive statistics of the data.

Using these data, we estimate the proposed concomitant function model with an increasing number of latent classes, beginning with two, until there is no improvement in the Bayes information criterion (BIC). We examine three different concomitant function models, with linear, quadratic, and cubic functions of the purchase frequencies. The optimal solution in our data is six latent classes with a quadratic function specification of the concomitant model. We then estimate the conventional model for six latent classes. The rationale for applying the model selection criterion to the proposed model rather than the conventional model is that we expect the conventional model to be misspecified, because it ignores the dependence between purchase frequency and choice parameters. For any given number of latent classes, the conventional model is nested within the proposed concomitant function model, and hence the likelihood ratio test may be used to compare these models. The quadratic concomitant function model has 69 parameters, compared with 59 parameters in the conventional model. The likelihood ratio test yields a chi-squared statistic of 104, which rejects the conventional model ( $p < .005$ ).

In addition to brand specific constants, we include in the deterministic component of brand utility the shelf price (in \$ per oz.) and indicator variables for whether or not the item was on special display or was feature-advertised. Table 5 gives results of maximum likelihood estimation of the six-segment conventional latent-class model, and Table 6 gives results for the proposed concomitant function model. Segment labels are matched between the tables by visually comparing the estimates. The signs and magnitudes of the estimated parameters of the proposed model are face-valid. Segments 2 and 3 have un-

intuitive (positive) signs of the price coefficient. However, the coefficient for segment 3 is not statistically different from 0. The positive coefficient of segment 2 is possibly explained by positive correlation between perceived quality and price among members of this segment. This explanation is consistent with their preference for Jif 28 oz., the highest price item.

### 5.1 Substantive Differences in Segmentation Structure

Comparing of Tables 5 and 6 demonstrates that the concomitant function model not only fits the data better, but also provides different substantive conclusions with respect to the segmentation structure. Note that the range of price coefficients is larger in the concomitant function solution. The most price-sensitive segment estimated by the concomitant model relative to the conventional approach is more price sensitive (price coefficient of  $-148$  vs.  $-126$ ). Further, the mean price coefficient estimate from the concomitant model is  $-71$ , versus  $-66$  from the conventional model.

An important caveat to note with respect to the empirical results is that estimated differences in preference and responsiveness parameters between segments may partly reflect differences in the (unidentified) scale parameter of the stochastic component of the utility (Swait and Louviere 1993; Swait 1994). This is the case with both the conventional and the proposed models. One might further expect that differences in the purchase frequencies between segments might be related to differences in the scale parameter. For example, households that purchase in the category more often may be more confident in their assessments of brands, resulting in lower variance of the utility for that segment.

Differences between the two models are also reflected in the estimated price elasticities from the models. In Table 7 gives

Table 5. Maximum Likelihood Estimates for Peanut Butter Data: Conventional Latent Class Logit Model

Parameter	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Jif 18 oz.	8.40	3.79	-2.50	.76*	2.92	6.18
Jif 28 oz.	1.18*	5.11	-1.94	-1.73*	1.41	3.88
Jif 40 oz.	-8.05	3.62*	-1.66	-14.28	-.92*	.09*
Peter Pan 18 oz.	11.71	1.75	-1.79	.18*	3.43	3.26
Skippy 18 oz.	9.84	1.65*	-1.00	3.79	3.33	3.31
Skippy 28 oz.	6.57	.97*	-.42	1.36	1.00	1.84*
Shelf price (\$/oz.)	-48.42	6.69	-9.28	-126.36	-94.48	-35.91
Display	2.69	-8.10	.59	.67*	1.21	.91*
Feature	1.88	1.72	2.06	.57*	2.06	1.80
Segment sizes	.077	.075	.184	.168	.387	.109

NOTE: Log-likelihood =  $-5,828.67$ . Skippy 40 oz. was treated as the base item. The  $p$  value of all parameter estimates is  $< .05$  except where marked with an asterisk.

Table 6. Maximum Likelihood Estimates for Peanut Butter Data: Concomitant Function (T and T<sup>2</sup>) Latent Class Logit Model

Parameter	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
Jif 18 oz.	-1.58	1.03	-3.71	1.54	2.87	6.16
Jif 28 oz.	-1.04	2.37	-4.21	-2.13*	1.39	3.93
Jif 40 oz.	-1.66	1.24	-2.80	-9.19	-.97	.05*
Peter Pan 18 oz.	-.80	-.44*	-2.66	.22*	3.85	3.21
Skippy 18 oz.	.49	-.78	-1.95	4.06	3.27	3.28
Skippy 28 oz.	.97	-1.01	-2.24	.17*	.88	1.88
Shelf price (\$/oz.)	-33.18	4.31	1.27*	-148.24	-81.77	-36.35
Display	.06*	.03*	1.59	1.07*	1.29	.87*
Feature	2.14	1.77	1.82	.74	2.01	1.77
Segment sizes	.133	.091	.058	.195	.422	.101

NOTE: Log-likelihood = -5,772.37. Skippy 40 oz. was treated as the base item. The p value of all parameter estimates is <0.05, except those marked with an asterisk. Coefficient estimates of the concomitant function are not shown.

estimated own-price and cross-price elasticities from the concomitant model, and the extent to which these deviate from estimated elasticities from the conventional latent-class model. The conventional model underestimates the own-price elasticities in magnitude quite substantially (with one exception, Skippy 18 oz.). Results on cross-elasticities are mixed.

### 5.2 Assignment of Panelists to Segments

We assign households to segments based on their posterior segment membership probabilities from the concomitant model. By examining the histogram of purchase frequencies, we find that the purchase frequencies varies considerably across segments. This observation validates the finding that in these data, the distribution of purchase frequency is dependent on segment membership.

Next we examine the amount of disparity in the assignments of panelists to the six segments from the conventional and the concomitant models. To quantify the dissonance between two cluster assignments, we create a 2 x 2 contingency table of a row variable and a column variable defined as follows: For each pair of panelists, the row variable has value 1 if the two panelists were assigned to the same segment by the conventional latent class approach and 0 otherwise. The column variable is defined similarly based on the segmentation produced by the concomitant function approach. The number of observations in

this contingency table is the number of distinct pairs of panelists. Therefore, because we have 829 panelists, the following contingency table has 343,206 observations:

	0	1
0	228,072	27,211
1	22,497	65,426

Clearly, the smaller the dissonance between the two sets of cluster assignments, the smaller the strength of the off-diagonal elements. The off-diagonal elements would be 0 under perfect agreement. We use the Hubert gamma statistic (Jain and Dubes 1988), which can be interpreted as the simple correlation between the row variable and the column variable, to quantify the strength of departures from perfect agreement. For these data, the Hubert gamma statistic is .627. Because the “observations” in the foregoing contingency table are not independent, we must resort to Efron’s (1979) bootstrap method to obtain confidence intervals. Using a bootstrap sample of 2,500, we estimate the 99% confidence interval to be .627 ± .0633, thereby rejecting the null hypothesis of perfect agreement between the segment assignment from the conventional latent-class model and the segment assignment from the concomitant function latent-class model. Our analysis therefore reveals that the two approaches have substantial differences in the assignments of panelists to clusters.

Table 7. Estimated Price Elasticities From Concomitant Model (percentage deviation from conventional model\*)

	Jif 18 oz.	Jif 28 oz.	Jif 40 oz.	Peter Pan 18 oz.	Skippy 18 oz.	Skippy 28 oz.	Skippy 40 oz.
Jif 18 oz.	-7.80 (27.6)	.87 (-4.0)	.86 (-4.0)	.81 (-5.1)	.83 (-3.7)	.86 (-2.4)	.84 (-4.5)
Jif 28 oz.	.13 (-17.1)	-8.68 (25.0)	.13 (-16.6)	.13 (-17.1)	.13 (-17.2)	.13 (-16.6)	.13 (-16.8)
Jif 40 oz.	.01 (-49.1)	.01 (-49.3)	-8.71 (24.5)	.01 (-48.5)	.01 (-48.2)	.01 (-48.6)	.01 (-48.9)
Peter Pan 18 oz.	1.75 (-.7)	1.79 (-.5)	1.79 (-.3)	-6.65 (30.6)	1.71 (-.4)	1.74 (-.1)	1.71 (-.7)
Skippy 18 oz.	5.36 (41.1)	5.44 (41.0)	5.38 (41.1)	5.17 (40.9)	-3.22 (-2.4)	5.17 (41.0)	5.29 (41.1)
Skippy 28 oz.	.41 (-6.6)	.42 (-6.6)	.42 (-6.4)	.40 (-6.6)	.41 (-6.6)	-8.04 (25.9)	.41 (-6.4)
Skippy 40 oz.	.14 (9.7)	.14 (9.2)	.14 (8.9)	.13 (9.3)	.14 (10.4)	.14 (9.3)	-8.39 (24.7)

\*Defined as  $\frac{\eta_{concomitant} - \eta_{conventional}}{\eta_{concomitant}} \times 100$ .

## 6. CONCLUDING COMMENTS

The random-coefficients logit model of conditional brand choice is firmly established in the marketing science literature. In this article we identify a key assumption underlying the extant approach to maximum likelihood estimation of the model parameters. This assumption is that the random-coefficients distribution of brand-choice parameters is independent of the purchase frequencies. We argue that this assumption is typically violated in commonly available household scanner panel data, resulting in biased and inconsistent estimates. In numerical simulations, we demonstrate that the bias can be quite severe.

We propose a conditional likelihood approach to obtain consistent estimates of brand-choice parameters when the dependence assumption may be violated in the data. We show that the proposed model can be expressed as a concomitant function latent-class model in which functions of purchase frequencies serve as concomitant variables. We contrast our proposed approach with the joint likelihood approach, in which the joint distribution of brand choice and purchase frequencies is estimated. A merit of our approach is that the distribution of purchase frequencies need not be fully specified to obtain unbiased estimates of the brand-choice parameters.

Future research should investigate the development of the conditional likelihood approach in the continuous random-coefficients model. We anticipate that, analogous to the current work, the solution may entail allowing the mean and variance-covariance matrix of the random-effects distribution to depend on purchase frequencies.

It is important to note that to implement our proposed solution, researchers need no more data than are currently available. Information on purchase frequencies of consumers is included in the panel data provided by companies like A. C. Nielsen and Information Resources Inc., and researchers need only modify the likelihood function.

In this article we are concerned with models of conditional brand choice alone. Another stream of work has modeled the brand choice and category purchase (incidence) decisions jointly (e.g., Chib et al. 2004; Bucklin and Gupta 1992). A natural question of interest is whether there is a similar source of bias in those models. Based on reasoning analogous to that offered in this article, one would expect that if the number of store visits is not independent of the parameters of the category purchase incidence model, then a similar bias should arise. We plan to investigate this conjecture more fully in future research.

## ACKNOWLEDGMENT

The comments of an associate editor and two reviewers are gratefully acknowledged.

[Received October 2003. Revised April 2004.]

## REFERENCES

- Bickel, P., and Docksum, K. (1977), *Mathematical Statistics*, Oakland, CA: Holden-Day.
- Brownstone, D., and Train, T. (1999), "Forecasting New Product Penetration and Flexible Substitution Patterns," *Journal of Econometrics*, 89, 109–129.
- Bucklin, R. E., and Gupta, S. (1992), "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach," *Journal of Marketing Research*, XXIX, 9, 201–215.
- Chiang, J. (1991), "A Simultaneous Approach to the Whether, What, and How Much to Buy Questions," *Marketing Science*, 10, 297–315.
- Chib, S., Seetharaman, P. B., and Strijnev, A. (2004), "Model of Brand Choice With a No-Purchase Option Calibrated to Scanner Panel Data," *Journal of Marketing Research*, to appear.
- Chintagunta, P. K. (1993), "Investigating Purchase Incidence, Brand Choice, and Purchase Incidence Decisions of Households," *Marketing Science*, 12, 184–209.
- Chintagunta, P. K., Jain, D. C., and Vilcassim, N. (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, XXVIII, 417–428.
- Chintagunta, P. K., and Prasad, A. (1998), "An Empirical Investigation of the Dynamic McFadden Model of Purchase Timing and Brand Choice: Implications for Market Structure," *Journal of Business & Economic Statistics*, 16, 2–12.
- Dayton, M. C., and Macready, G. B. (1988), "Concomitant-Variable Latent-Class Models," *Journal of the American Statistical Association*, 83, 173–179.
- Dillon, W., and Gupta, S. (1996), "A Segment-Level Model of Category Volume and Brand Choice," *Marketing Science*, 15, 38–59.
- Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Applied Statistics*, 7, 1–26.
- Fader, P., and Hardie, B. (1996), "Modeling Consumer Choice Among SKU's," *Journal of Marketing Research*, XXXIII, 442–452.
- Forsyth, J., Gupta, S., Haldar, S., Kaul, A., and Kettle, K. (1999), "A Segmentation You Can Act on," *The McKinsey Quarterly*, 3, 6–15.
- Gonul, F., and Srinivasan, K. (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science*, 12, 213–229.
- Gupta, S., and Chintagunta, P. K. (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models," *Journal of Marketing Research*, XXXI, 128–136.
- Jain, A., and Dubes, R. (1988), *Algorithms for Clustering Data*, Englewood Cliffs, NJ: Prentice-Hall.
- Kamakura, W., and Russell, G. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, XXVI, 379–390.
- Kim, B.-D., and Rossi, P. E. (1994), "Purchase Frequency, Sample Selection, and Price Sensitivity: The Heavy User Bias," *Marketing Letters*, 5, 57–67.
- Mood, A., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, New York: McGraw-Hill.
- Swait, J. (1994), "A Structural Equation Model of Latent Segmentation and Product Choice for Cross-Sectional Revealed Preference Choice Data," *Journal of Retailing and Consumer Services*, 1, 77–89.
- Swait, J., and Louviere, J. (1993), "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, XXX, 305–314.