

ACCOUNTING FOR HETEROGENEITY AND NONSTATIONARITY IN A CROSS-SECTIONAL MODEL OF CONSUMER PURCHASE BEHAVIOR

PETER S. FADER AND JAMES M. LATTIN

*University of Pennsylvania
Stanford University*

When calibrating a brand choice model cross-sectionally, a measure of brand loyalty is often introduced into the utility function to account for differences in utility across households and over time. One of the most widely used measures of brand loyalty, proposed by Guadagni and Little (1983), is an exponential smoothing model of past choice behavior by the household. In this study, we argue that the exponential smoothing model of brand loyalty cannot properly distinguish between sources of variation in utility due to heterogeneity (across households) and sources of variation due to nonstationarity (within household over time). We introduce a new measure of brand loyalty, derived from a nonstationary Dirichlet-multinomial choice model, in which heterogeneity and nonstationarity are handled distinctly.

(Choice Models; Brand Loyalty; Heterogeneity; Nonstationarity)

1. Introduction

Many researchers have examined brand choice behavior using multinomial logit models calibrated on scanner panel data. These cross-sectional models generally allow for brand utilities that potentially differ across households and over time. Researchers have used both fixed effects (e.g., Jones and Landwehr 1988) and random effects (e.g., Chintagunta et al. 1991) specifications to capture the cross-sectional variation. While these methods are useful in explaining the differences in brand preferences across households, they do not allow for any changes in consumer tastes over time.

A popular approach that captures *both* types of variation is the exponential smoothing model of brand loyalty proposed by Guadagni and Little (1983). This measure tracks changes in household tastes over time using a weighted average of past choice behavior in which recent choice is weighted more heavily. Thus, this loyalty variable captures "not only much of the cross-sectional heterogeneity but also a good part of the purchase-to-purchase dynamics" (Guadagni and Little 1983, p. 216). Unfortunately, while this measure embodies both sources of variation, it is unable to separate out these two effects. One cannot tell whether recent choice behavior improves the fit of the choice model because it increases the precision of the household-level loyalty estimates or because it tracks the changes in utility over time. As shown by Massy et al. (1970), ignoring cross-sectional heterogeneity may give the appearance of nonstationarity in preference when in fact none exists. This confusion is a problem when testing theories related to factors

affecting the household that change over time (see Neslin and Shoemaker 1989 and Ortmeyer et al. 1991).

To address this concern, we present a new measure of brand loyalty to sort out heterogeneity and nonstationarity in cross-sectional choice models. We derive our measure from a nonstationary Dirichlet-multinomial model of brand choice behavior, similar to the beta-binomial-geometric model proposed by Sabavala and Morrison (1981). Our approach recognizes that the observed choice behavior of the household enables us to do two things: (1) update our priors regarding the probabilities governing the choice behavior of any particular household, and (2) track changes in these probabilities over time. Even when the probabilities governing household choice behavior do not change over time, we can still use the observed choice behavior to update our priors. With the exponential smoothing model of brand loyalty, the only way to accomplish (1) is to force the occurrence of (2). Our approach provides for a single measure of brand loyalty in which (1) and (2) are handled distinctly.

The paper is organized as follows. In § 2, we review the use of the exponential smoothing model and in § 3 we derive our measure of brand loyalty. In § 4, we use simulation to demonstrate the limitations of the exponential smoothing model of brand loyalty; using the same simulation data, we then demonstrate the advantages of the proposed loyalty measure. In § 5, we compare and contrast the results of our approach to existing models using actual scanner panel data on refrigerated ready-to-drink orange juice.

2. Brand Loyalty: Exponential Smoothing

Following the precedent established by Guadagni and Little (1983), most recent research into conditional brand choice behavior uses a multinomial logit model (MNL) with the following structure:

$$\text{MNL}_{ht}(i) = \frac{\exp(v_{ht}(i))}{\sum_k \exp(v_{ht}(k))} \quad \text{where} \quad (1)$$

$\text{MNL}_{ht}(i)$ = the probability that household h chooses brand i on purchase occasion t ,

$v_{ht}(i) = \alpha_i + \gamma \text{LOY}_{ht}(i) + \sum_m \beta_m x_{mt}(i)$ = the utility to household h of purchasing brand i on occasion t ,

$\text{LOY}_{ht}(i)$ = household h 's loyalty to brand i on occasion t , based on purchases before time t ,

$x_{mt}(i)$ = the level of marketing mix variable m for brand i on occasion t , and

$\alpha_i, \gamma, \beta_m$ = model parameters, assumed to be constant over time and across households.

The exponentially smoothed loyalty measure (hereafter called SMOOTH) takes the following form:

$$\text{SMOOTH}_{ht}(i) = \lambda \text{SMOOTH}_{ht-1}(i) + (1 - \lambda)y_{ht-1}(i) \quad \text{where} \quad (2)$$

$y_{ht}(i) = 1$ if household h purchased brand i on occasion t and 0 otherwise, and

λ = smoothing parameter, assumed to be constant over time and across households.

The exponential smoothing model can be derived as a weighted average of past purchases under certain assumptions regarding the steady-state nature of the measure (see Srinivasan and Kesavan 1976).

Most researchers use each household's early purchase history from to "start up" the brand loyalty measure; Guadagni and Little (1983), for example, set aside the first 25 weeks of their study as an initialization period. On the first purchase occasion for each household, in the absence of any information on prior purchase behavior, we set $\text{SMOOTH}_{h1}(i)$ equal to $1/n$ for all brands i , where n is the total number of brands in the analysis. To calculate SMOOTH, one also needs to know the value of the smoothing

parameter λ ; maximum likelihood estimates of this parameter can be obtained through a Taylor series expansion of SMOOTH (see Fader et al. 1992). In most applications, λ falls into the range between 0.70 and 0.90.

2.1. *Pitfalls*

The SMOOTH measure is not without its limitations in cross-sectional models of consumer choice. One problem is the assumption that SMOOTH is in “steady state.” Despite having as much as one year of data for initializing the measure, in many applications this enables us to see only one or two purchases by light users. For those households with short choice histories, our priors will play a relatively more important role in determining choice probabilities than for households with long choice histories.

Problems of interpretation may also arise. A value of λ less than one is not conclusive evidence of nonstationarity. This is because SMOOTH is based on past choice behavior; i.e., a relatively small sample of choice outcomes from a probabilistic process. By updating SMOOTH after each choice made by the household, we increase the amount of sample information used to calculate the measure and thereby decrease its error variance. Only when $\lambda < 1$ is the exponential smoothing model able to utilize the entire purchase history of the household in SMOOTH, which gives the appearance of nonstationarity.

One consequence of using SMOOTH when choice behavior is truly stationary is that the optimal value of the smoothing parameter λ will depend upon the average length of the choice history of the household. If the choice history is very short (e.g., two or three purchases), then λ will be considerably less than one. Otherwise, if λ is too close to one, the weight on the most recent choices will be negligible compared to the weight on the values used to initialize SMOOTH. If the choice history is very long, then λ will approach 1; otherwise, the weights on the observations from the early choice history will quickly go to zero and be lost. This result suggests that in choice models using SMOOTH, the estimate of the smoothing parameter will not be independent of the amount of data used to calibrate the model. When choice behavior is in fact stationary, there will be a direct relationship between the length of the study (initialization period and calibration period) and the magnitude of λ .

3. Brand Loyalty: Nonstationary Dirichlet-Multinomial

We now propose a new measure of brand loyalty, which we derive from a stochastic model of choice behavior in the absence of marketing mix effects.¹ We assume Dirichlet heterogeneity for multinomial choice probabilities across households and use a Bernoulli distribution to describe the likelihood of a shift in choice probabilities between purchase occasions for a given household. We refer to this model as a nonstationary Dirichlet-multinomial (NSDM); when the Bernoulli probability of a change in choice probabilities is zero, the model becomes a stationary Dirichlet-multinomial model (SDM). We derive the expected brand choice probabilities given the observed choice behavior of the household. These posterior probabilities form the basis of our proposed measure of brand loyalty.

3.1. *The NSDM Model*

We begin by stating the assumptions underlying the NSDM model. First, we assume Dirichlet heterogeneity in choice behavior across households. Thus, for a given household, our priors about the multinomial probabilities governing choice from among n brands are described by a Dirichlet distribution with parameters $(\tau_1, \tau_2, \dots, \tau_n)$, i.e., a distribution function with pdf

¹ In § 5, where we test the measure using actual purchase panel data, we modify the choice model to accommodate the effects of point-of-purchase marketing mix variables such as price and promotion.

$$D(p_1, p_2, \dots, p_n) = \frac{\Gamma(\tau_1 + \tau_2 + \dots + \tau_n)}{\Gamma(\tau_1)\Gamma(\tau_2) \dots \Gamma(\tau_n)} p_1^{\tau_1-1} p_2^{\tau_2-1} \dots p_n^{\tau_n-1} \quad (3)$$

where $\sum_i p_i = 1$. The use of the Dirichlet to capture cross-sectional heterogeneity in choice behavior for frequently purchased packaged goods is well established (e.g., Ehrenberg 1988, Fader 1993).

Second, we assume that the nonstationarity in choice behavior of a given household is described by a renewal process. The probability that the choice probabilities for a given household are renewed between choice occasions (where renewal constitutes a new draw from the original Dirichlet distribution, independent of previous choice behavior) is Bernoulli with probability $(1 - \lambda)$, and that the renewal rate is the same across households. Thus, for any household the number of choice occasions since the last renewal is geometrically distributed with pdf

$$G(x) = (1 - \lambda)\lambda^x \quad \text{where} \quad x = 0, 1, 2, 3, \dots \quad \text{and} \quad 0 \leq \lambda \leq 1. \quad (4)$$

There are many sudden, unforeseen occurrences (even in relatively mature product categories) that might cause the kind of abrupt change in household choice probabilities suggested by this renewal process. Examples include the entry or exit of a brand, a price war or some unusual promotional activity (e.g., a new advertising campaign). These major changes are the most important to capture; more gradual changes in underlying preference may be so slight that a stationary Dirichlet model will hold reasonably well over a long period of time (Ehrenberg 1988, p. 12).

In behavioral terms, our model stands in sharp contrast to SMOOTH, which implicitly assumes a more gradual change in consumer choice probabilities. However, as we shall show, both models lead to measures that allow for the gradual updating of brand loyalty as a function of past choice behavior. Through simulation and empirical testing, we hope to establish which of the two is a better representation based on fit and predictive ability.

The two assumptions above yield the NSDM model, which is a multivariate extension of the beta-binomial-geometric model used by Sabavala and Morrison (1981) to describe the frequency distribution of exposures to a particular media vehicle (such as TV or magazines).² Our goal is to use the NSDM model and the observed past choice history of the household through time t to derive the posterior choice probabilities. These posteriors represent our best estimates of the probabilities governing the choice behavior of the household on occasion $t + 1$. We denote the purchase history of household h through time t by $y_h = (y_{h1}, y_{h2}, \dots, y_{ht})$, where $y_{h1} = (y_{h1}(1), y_{h1}(2), \dots, y_{h1}(n))$ is a vector of $(0, 1)$ variables indicating which brand was purchased on the household's first purchase occasion.

We begin by finding the expected probabilities conditional on the number of choice occasions since the last renewal. For example, if the number of choice occasions since the last renewal is x , then we know that the choice behavior on occasions $t, t - 1, \dots, t - x + 1$ was governed by the same set of multinomial choice probabilities. Combining our priors with the x multinomial choice outcomes indicated by $y_{ht}, y_{ht-1}, \dots, y_{ht-x+1}$ yields the posterior distribution. Because the Dirichlet and the multinomial form a conjugate pair, the posterior distribution is also Dirichlet with parameters

$$(\tau_1 + \sum_c y_{hc}(1), \tau_2 + \sum_c y_{hc}(2), \dots, \tau_n + \sum_c y_{hc}(n)),$$

² Vanhonacker and Winer (1990) also develop a Dirichlet-multinomial geometric model of choice behavior. However, instead of using the Dirichlet to characterize heterogeneity across consumers (in fact, their model is calibrated at the individual level), they derive the Dirichlet as the optimal decision distribution assuming rational random behavior.

where c is a subscript of choice occasions from $t - x + 1$ to t . Thus, for household h with choice history y_h , the expected probability of choosing brand i on occasion $t + 1$ given x choice occasions since the last renewal is given by

$$E(p_i | y_h, x) = \frac{\tau_i + \sum_c y_{hc}(i)}{\tau_1 + \tau_2 + \dots + \tau_n + x} \quad (5)$$

where c subscripts choice occasions $t - x + 1$ to t .

We cannot tell exactly when a renewal of brand choice probabilities takes place. However, we can use the geometric distribution to calculate $E(p_i | y_h)$, the expected choice probabilities for occasion $t + 1$ unconditional on the length of time since the last renewal:

$$E(p_i | y_h) = \sum_x E(p_i | y_h, x) G(x) + \lambda^t E(p_i | y_h, t), \quad (6)$$

where x indexes the number of choice occasions since the last renewal (and ranges from 0 to $t - 1$). Note that in the second term on the right-hand side of the equation, $\lambda^t = 1 - \sum_x G(x)$. For example, when $n = 2$ and $t = 2$, $E(p_i | y_h)$ is written as

$$(1 - \lambda) \frac{\tau_1}{\tau_1 + \tau_2} + \lambda(1 - \lambda) \frac{\tau_1 + y_{h2}(1)}{\tau_1 + \tau_2 + 1} + \lambda^2 \frac{\tau_1 + y_{h1}(1) + y_{h2}(1)}{\tau_1 + \tau_2 + 2}. \quad (7)$$

The expression in (7) reveals that when choice behavior is nonstationary, we weight the most recent choice information most heavily in determining $E(p_i | y_h)$. The more nonstationary the process (i.e., the greater the likelihood of a renewal between choice occasions), the more rapidly we discount the information from the early choice history of the household. When choice behavior is purely stationary, the information from each of the past choice occasions is equally weighted in determining $E(p_i | y_h)$.

Our proposed measure of loyalty is based on $E(p_i | y_h)$. In order to put it into the context of the cross-sectional multinomial logit model, we set

$$\text{NSDM}_{ht}(i) = \ln [E(p_i | y_{h1}, y_{h2}, \dots, y_{ht-1})], \quad (8)$$

where $\text{NSDM}_{ht}(i)$ is our measure of loyalty to brand i by household h at time t . In the absence of marketing mix effects, if we set $v_{ht}(i) = \text{NSDM}_{ht}(i)$ and substitute into Equation (1), the logit choice probability $\text{MNL}_{ht}(i)$ is exactly the expected choice probability, $E(p_i | y_h)$.

In order to calculate $\text{NSDM}_{ht}(i)$ for each brand i , we need information on the ordered purchase history of household h through purchase occasion $t - 1$. This is exactly the same information necessary to calculate $\text{SMOOTH}_{ht}(i)$, which is easily available from the scanner panel dataset.

3.2. NSDM vs. SMOOTH

Our measure NSDM differs from the exponential smoothing model SMOOTH in two respects. First, our approach involves a different characterization of the nonstationarity in brand choice behavior. SMOOTH is based on a linear-learning model, in which the choice probabilities of the household move in deterministic fashion in the direction of the choice made on the latest purchase occasion. In contrast, NSDM is based on a renewal model, in which the new choice probabilities are completely independent of past values. This specification enables us to make inferences about the nonstationarity of brand choice behavior that are not confounded by heterogeneity.

A second point of difference between NSDM and SMOOTH is our explicit treatment of cross-sectional heterogeneity using the Dirichlet distribution. The Dirichlet parameters $(\tau_1, \tau_2, \dots, \tau_n)$ reflect our priors about the probabilities governing the choice behavior of each household. Thus, NSDM takes into account the difference in frequency of choice across households in a way that SMOOTH cannot. For a light user, the cumulative

impact of the short prior purchase history is relatively small; however, the impact of each additional choice observation in updating NSDM (in the absence of a long purchase history) is relatively large. For a heavy user, we are likely to get more extreme values of NSDM, especially when the household favors one or two brands over the alternatives.

The Dirichlet parameters also provide evidence about the level of heterogeneity in the market. The summary statistic $\phi = 1/(1 + \sum_i \tau_i)$, used by numerous researchers (e.g., Jeuland et al. 1980; Sabavala and Morrison 1977), is a well-accepted measure of heterogeneity. A value of ϕ approaching zero indicates complete homogeneity, while a value of ϕ approaching one signifies extremely high heterogeneity across households.

4. Simulation

Using simulated data, we now compare the performance of the two loyalty measures SMOOTH and NSDM. We create a “world” in which there are no marketing mix effects and in which choice behavior is perfectly stationary. We then calibrate two cross-sectional choice models: one based on SMOOTH and the other based on NSDM. For the SMOOTH model, we show that the estimated value of the parameter λ is substantially less than 1.0 and that the bias in λ is directly related to the amount of data used to measure brand loyalty. For the NSDM model, we show that the estimate of λ has no substantial bias, that its magnitude is not related to the amount of data used in the measure, and that the fit of the model improves significantly.

4.1. Data

In order to make the simulation realistic, we derive its parameters from actual purchase data. The scanner data (provided by Information Resources, Inc.) describe consumer purchase behavior of refrigerated orange juice in Marion, Indiana, over a two-and-a-half year period (January 1983 to June 1985). We selected a random sample of 200 households from among all households who purchased at least one of the following six leading brands (in the 64 oz. size) during 1984: Citrus Hill, Minute Maid, Tropicana Regular, Tropicana Premium, a regional brand and a private label brand. During the first two years of the sample period, our sample households made a total of 3079 purchases.³

For each household, we calculate the choice share of each brand over the first two years of data. Treating these household-level choice shares as multinomial probabilities, we use random numbers to simulate purely stationary purchase behavior by the household: one simulated choice for each actual purchase made by the household during the two years. In this way, we insure that the simulation embodies an empirically-grounded distribution of brand preferences and purchase frequencies across households. For each of the two models (SMOOTH and NSDM), we use the second year of simulated data (i.e., the choices corresponding to the purchase occasions in weeks 53 to 104) to calibrate the model parameters. To test our conjectures regarding the relationship between the amount of data and the bias in the estimate of λ , we vary the length of the initialization period used to “start up” SMOOTH and NSDM. We use four different lengths: zero weeks, 13 weeks, 26 weeks, and 52 weeks. Our hypothesis suggests that for SMOOTH, the estimated value of λ should get closer to one as the initialization period increases in length; for NSDM, the estimated value of λ should be close to one in all cases.

4.2. Simulation Results

We ran this simulation 50 times. For each run, we performed eight model calibrations: each of the two models (SMOOTH and NSDM) for each of the four different initialization periods. Figure 1 shows the average model fit (indicated by the log likelihood value)

³ We hold out the last six months of data (during which time the panelists made 666 purchases) for the purposes of model validation in § 5 of the paper.

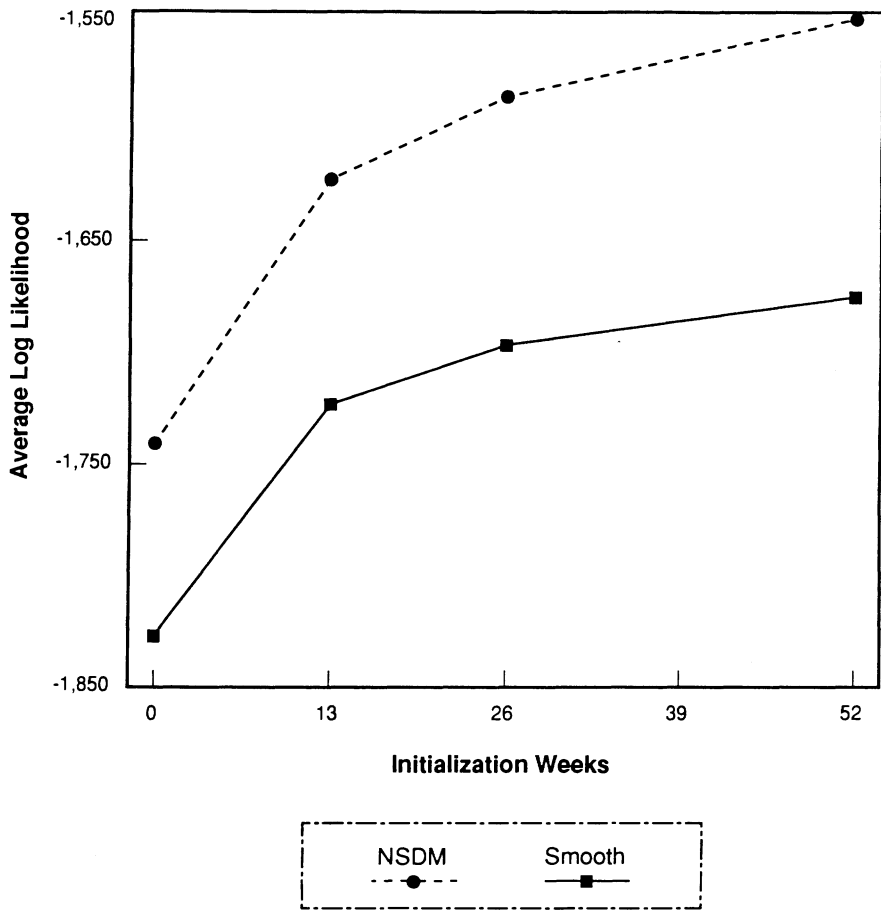


FIGURE 1. Simulation Results: Average Fit (Log Likelihood) Across Models and Across Initialization Periods.

across all 50 runs of the simulation for each of the eight cases. Two noteworthy patterns are evident. First, regardless of the underlying model, an increase in the length of the initialization period leads to better model fits.⁴ This is consistent with our expectations: the more data available to establish the brand loyalty measure, the lower its error variance. Second, holding the length of the initialization period constant, NSDM shows a significant improvement in fit ($p < 0.001$) over SMOOTH for each of the four different initialization periods.

Figure 2 shows the average estimate of the smoothing parameter λ for the models SMOOTH and NSDM for each of the four initialization periods. We note three things. First, for the SMOOTH model, the average estimated value of λ is substantially and significantly less than 1.0 in all cases. Second, for the same model, the average estimated value of λ increases significantly with each increase in the length of the initialization period. Thus, as conjectured, the bias in λ does seem to depend upon the amount of data used to establish the measure SMOOTH. Third, for the NSDM model, the average estimate of the smoothing parameter λ is in all cases very close to one. The fact that the NSDM smoothing parameter is seemingly independent of the length of the data period makes the model an attractive option for use across datasets of different sizes and different product categories.

⁴ For each model, the differences in fit as the initialization period increases in length are all significant at $p < 0.001$.

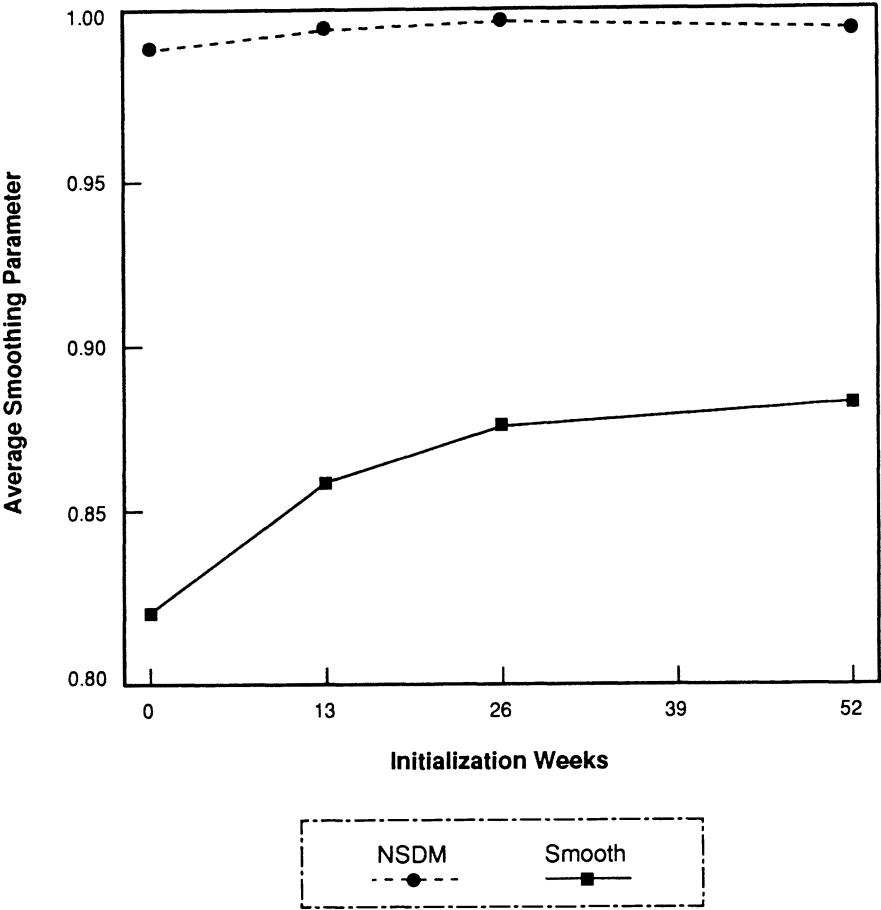


FIGURE 2. Simulation Results: Average Smoothing Parameter Estimate Across Models and Across Initialization Periods.

5. Application: Refrigerated Orange Juice

The simulation results suggest that there are circumstances—when choice behavior is stationary and marketing mix effects are absent—in which the proposed measure outperforms the exponential smoothing model of brand loyalty in a cross-sectional model of brand choice behavior. It remains for us to show, however, that the superior performance of the proposed model holds up in a “real-world” choice situation: one where household choice behavior may not be stationary and in which there are pronounced marketing mix effects. We now test the proposed approach using the actual IRI scanner panel data on refrigerated ready-to-drink orange juice.

5.1. Models

As in the simulation, we compare the performance of the exponential smoothing model of brand loyalty (SMOOTH) with the proposed nonstationary Dirichlet-multinomial model (NSDM). The complete specification of the utility function for the model based on SMOOTH is repeated below:

$$v_{ht}(i) = \alpha_i + \gamma \text{SMOOTH}_{ht}(i) + \sum_m \beta_m x_{mt}(i) \tag{9}$$

where the measure SMOOTH is as defined in Equation (2). Similarly, we modify the utility function of the NSDM model to take into account marketing mix effects as follows:

$$v_{ht}(i) = \alpha_i + \text{NSDM}_{ht}(i) + \sum_m \beta_m x_{mt}(i) \quad (10)$$

where the measure NSDM is as defined in Equation (8). We include the brand-specific constants to account for the effects of possible omitted variables (such as product quality and media advertising). Including these intercept terms means that NSDM will require $(n - 1)$ parameters more than SMOOTH, where n is the number of available brands.

In an actual purchase setting, price and promotion directly influence household choice behavior. Although we have assumed in our derivation that y_{hc} reflects only the underlying brand choice probabilities, the observed choice behavior of each household in fact reflects not only household tastes but also the impact of time-varying marketing mix variables. Thus, NSDM no longer serves as a stand-alone choice model, but as a loyalty variable within a multinomial logit model. Nonetheless, we feel that the derivation in § 3 provides a strong intuitive rationale for the measure NSDM and that all of the parameter interpretations discussed earlier still apply. Whether or not the proposed approach performs well in an actual choice setting should be resolved empirically.

We also assess the performance of a special case of NSDM in which the nonstationarity parameter λ is constrained to 1. This model, which we denote SDM (for stationary Dirichlet-multinomial), captures only the heterogeneity in choice behavior across households. The form for the utility function is exactly the same as for NSDM except that the brand loyalty measure simplifies to:

$$\text{SDM}_{ht}(i) = \ln \left[\frac{\tau_i + \sum_c y_{hc}(i)}{\tau_1 + \tau_2 + \cdots + \tau_n + t - 1} \right] \quad (11)$$

where c subscripts the choice occasions from 1 to $t - 1$. By comparing the performance of SDM with NSDM, we can assess the incremental benefit of tracking nonstationarity.

We compare NSDM with two other models designed to account for cross-sectional heterogeneity in choice behavior. The first is a multinomial logit model which uses as its measure of brand loyalty the “share of previous purchase” variable proposed by Krishnamurthi and Raj (1988):

$$\text{SHARE}_{ht}(i) = \frac{\sum_c y_{hc}(i)}{\sum_c \sum_k y_{hc}(k)} \quad (12)$$

SHARE differs from SDM in two respects. First, it lacks the brand-specific parameters τ_i ; second, because SHARE can take on a value of zero, it cannot employ the log transformation used in SDM.

We also examine the semiparametric random-effects model proposed by Chintagunta et al. (1991). Instead of using a loyalty-type measure to capture cross-sectional heterogeneity, this model assumes that the fixed term (i.e., the brand-specific constant) in the utility function for each brand varies across households according to some underlying probability distribution. The semiparametric specification involves approximating each of these distributions using a finite number of support points, and then estimating the location and the probability mass associated with each support point.

More formally, let $f_i(\alpha_i)$ denote the probability density function describing the underlying distribution of α_i . In the semiparametric specification, one approximates $f_i(\alpha_i)$ with a finite number of support points, N_i . Chintagunta et al. (1991) find that three support points are sufficient to approximate the underlying distribution $f_i(\alpha_i)$. Therefore, we also set $N_i = 3$ for each brand i . Note that for the six brands in our orange juice data,

this relatively small number of support points results in a very large number of support point combinations ($3^6 = 729$).

5.2. Data

As described in § 4.1 above, our database consists of 200 panelists, 6 brands, and 3,079 purchases. All 52 weeks from the year 1983 (a total of 1,490 purchases) are treated as the sole initialization period, all of the data from 1984 (1,589 purchases) are used for calibration purposes, and 666 purchases from the first six months of 1985 are reserved for a predictive validation. Besides the different loyalty measures, all of our models include variables for regular (depromoted) price, short-term price cuts, and a 0-1 variable indicating the presence of any newspaper feature activity.⁵

Each of the choice models (except the semiparametric random-effects model) is calibrated as an ordinary multinomial logit model using an iterative Taylor series procedure (Fader et al. 1992). We calibrate the semiparametric model using the general purpose nonlinear optimization software GQOPT (Quandt and Goldfeld 1987). Because of the large number of support point combinations, this last procedure is nearly 1000 times more computationally intensive than the other four.

5.3. Results and Discussion

We initialize the four measures SHARE, SMOOTH, SDM and NSDM using the data from the initialization period. We then estimate the parameters for all five models using the data from the calibration period. The parameter estimates and model fit statistics are presented in Table 1 (with additional information on the semiparametric model included in Table 2). In terms of model fit, the NSDM model ($LL = -1401$, $\bar{\rho}^2 = 0.430$) comes out on top. Because the models are not nested, we use a test statistic from Ben-Akiva and Lerman (1985, p. 172) that transforms the difference in (non-nested) model fit to a normal random variable. In comparing the NSDM and SMOOTH models, the fit statistics from Table 1 yield a standard normal score of -8.69 , indicating that the difference is clearly significant ($p < 0.001$). The difference between NSDM with SDM is even more significant, which clearly conveys the importance of accounting for nonstationarity in this particular dataset.

We use the data from the holdout period to see whether the improvement in fit offered by the NSDM model holds up in a predictive test. Using the parameter estimates from Table 1 to predict purchase behavior in the holdout period, we find that NSDM ($LL = -602$, $\rho^2 = 0.389$) once again outperforms all four competing models.

Despite the limitations of SMOOTH illustrated in our simulation, its ability to fit and predict “real world” brand choice behavior is quite strong (and well established in the literature). The stationary models (SDM, SHARE and the semiparametric random-effects model) fall short in their ability to fit brand choice behavior apparently because they cannot account for nonstationarity. However, both SDM and the semiparametric model predict choice behavior in the holdout sample at least as well as SMOOTH. This suggests that the predictive performance of SMOOTH may be almost entirely due to its ability to account for the heterogeneity in household tastes.

The coefficients for the marketing mix variables (feature, regular price, and price cut) are all significant, intuitively reasonable, and fairly similar across the five models. The key difference is in the estimated value of λ . As in the simulation, the estimate of the nonstationarity parameter is substantially lower for SMOOTH (where we anticipate a downward bias) than for NSDM (where we anticipate no bias). In NSDM, the estimate of $\lambda = 0.93$ suggests that, on average, a renewal occurs between choice occasions with

⁵ Because the product must be refrigerated, there is virtually no display activity in the category.

TABLE 1
Fit Results and Estimated Model Coefficients: Refrigerated Orange Juice

		Nonstationary Models		Stationary Models		
		NSDM	Smooth	SDM	Semiparametric	SHARE
Brand-specific constants:						
Regional	α_1	-0.205 (-0.9) ^a	0.219 (1.2)	-0.207 (-0.9)	^d	0.131 (0.7)
Citrus Hill	α_2	0.656 (2.8)	1.161 (5.8)	0.886 (4.1)	^d	1.187 (6.3)
Tropicana Reg.	α_3	-0.187 (-0.8)	0.404 (2.0)	-0.092 (-0.4)	^d	0.376 (2.1)
Tropicana Prem.	α_4	1.478 (4.4)	0.884 (3.0)	1.236 (3.8)	^d	0.912 (3.2)
Minute Maid	α_5	0.636 (2.4)	1.133 (4.7)	0.852 (3.3)	^d	1.164 (5.1)
Private Label	—	0.000 ^b	0.000 ^b	0.000 ^b	^d	0.000 ^b
Loyalty	γ		3.927 (23.6)			2.945 (28.8)
Marketing mix coefficients:						
Feature	β_1	0.603 (6.1)	0.606 (6.0)	0.596 (6.0)	0.663 (5.8)	0.618 (6.3)
Regular Price	β_2	-2.625 (-10.3)	-2.482 (-9.6)	-2.775 (-11.0)	-3.299 (-10.4)	-2.675 (-10.8)
Price Cut	β_3	2.276 (9.2)	2.343 (9.3)	2.287 (9.3)	2.685 (9.3)	2.360 (9.7)
Brand-loyalty coefficients:						
Regional	τ_1	0.240 (3.6)		0.760 (4.3)		
Citrus Hill	τ_2	0.248 (4.4)		0.666 (5.0)		
Tropicana Reg.	τ_3	0.303 (3.5)		0.880 (4.0)		
Tropicana Prem.	τ_4	0.024 (2.2)		0.117 (2.3)		
Minute Maid	τ_5	0.269 (4.0)		0.751 (4.5)		
Private Label	τ_6	0.081 (3.0)		0.242 (3.1)		
Nonstationary	λ	0.926 (-6.7) ^c	0.831 (-12.1) ^c	1.000 ^b		
Calibration Period ($N = 1589$):						
Log Likelihood:		-1401	-1438	-1458	-1503	-1523
Parameters:		15	10	14	32	9
Fit statistic	$\bar{\rho}^2$	0.430	0.417	0.407	0.382	0.383
Prediction Period ($N = 666$):						
Log likelihood:		-602	-621	-621	-619	-662
Fit statistic	ρ^2	0.389	0.370	0.370	0.372	0.328

^a t-statistics in parentheses.
^b Value of the parameter constrained in estimation.
^c reported t-test versus null hypothesis $\lambda = 1$.
^d See Table 2.

TABLE 2
Parameter Estimates from Semiparametric Model

Brand	First Support Point		Second Support Point		Third Support Point		Expected Value
	Location	Prob. Mass	Location	Prob. Mass	Location	Prob. Mass	
Regional	1.679	0.086	-1.406	0.613	-2.031	0.301	-1.327
Citrus Hill	2.332	0.232	0.298	0.313	-1.394 ^a	0.455	0.000
Tropicana Regular	0.753	0.229	-1.250	0.625	-21.541	0.146	-3.755
Tropicana Premium	2.435	0.118	-0.873	0.189	-21.486	0.692	-14.755
Minute Maid	23.017	0.025	0.831	0.583	-1.552	0.392	0.452
Private Label	4.310	0.106	-0.343	0.347	-5.688	0.548	-2.779

^a This parameter restricted so that the expected value of the distribution for Citrus Hill equals zero.
Each row in this table describes the semiparametric probability distribution function for each brand. Each brand's pdf is characterized by three *support points*, and each support point has an associated probability mass. These probabilities can be interpreted as the proportion of households whose intrinsic preference for the brand is equal to the corresponding support point location.
Consider the regional brand, for example. The support points and associated probability masses suggest that 8.6% of the sample households have a brand-specific intercept of 1.679, 61.3% have a brand-specific intercept of -1.406, and the remaining 30.0% have an intercept of -2.031 for the regional brand.
For further background on this model, see Chintagunta et al. (1991).

probability 0.07, or on average once every 14 purchases. To put this in perspective, the average panelist makes about 15 purchases over the initialization and calibration periods. Despite the fact that these renewal events occur rather infrequently, the highly significant difference between NSDM and SDM suggests that we still make great gains by capturing these occasional changes.

As discussed by Guadagni and Little (1983, p. 214), the brand-specific constants in the logit model “serve to capture any uniqueness an alternative has that is not captured by other explanatory variables.” To the extent that an omitted variable such as product quality is reflected in the average price of the brand, we might expect to see a high correlation between the brand-specific constants and average shelf price.⁶ In fact, the Spearman rank correlation between the estimates of α_i and the average shelf price of each brand is the same (0.71) for the SMOOTH and NSDM models. The difference between the two models is that while the brand-specific constants from SMOOTH are highly correlated with the brand market shares (0.60), the brand-specific constants from NSDM are not (−0.09).

Unlike the brand-specific constants, the estimates of the parameters τ_i in NSDM are reasonably independent of both the brand-specific constants and the average shelf prices (the rank correlations are −0.37 and −0.20, respectively). Apparently, these parameters are capturing a different phenomenon—which we believe to be more closely related to brand preferences—than the brand-specific constants.

The parameters τ_i also reveal the level of heterogeneity in the market. Based on the estimates in Table 1, $\phi = 0.46$, indicating a fairly substantial amount of heterogeneity. In contrast, the level of heterogeneity implied by the SDM, in which the parameter λ is constrained to 1, is much lower ($\phi = 0.23$). We expect this finding to hold in general; that is, by explicitly accounting for nonstationarity in purchasing patterns, household brand loyalties will appear to be less similar cross-sectionally than they would in a model that assumes pure stationarity.

6. Conclusion

In this paper, we have developed a new loyalty measure, derived from a nonstationary Dirichlet-multinomial model of brand choice, that is able to separate nonstationarity from heterogeneity in consumer purchase behavior. One of the benefits of our proposed model is that it requires only a few more parameters than the traditional exponential smoothing model (Guadagni and Little 1983), but it can be estimated with the same data and same multinomial logit software (using the estimation procedure proposed by Fader et al., 1992).

We demonstrated the potential value of the proposed approach using simulated and actual data. The logit model featuring the proposed measure of loyalty fit the actual data significantly better than existing models, including the logit model proposed by Guadagni and Little (1983); furthermore, this improvement in fit held up to predictive validation. The additional information provided by the estimates of the Dirichlet parameters τ_i is also potentially quite meaningful. These parameters not only reveal the level of heterogeneity in the market, but also tell us something about brand loyalty in the absence of observed choice behavior. We believe that these advantages make the proposed measure worthy of consideration in future modeling efforts.

⁶ The average shelf price for each brand is based upon the net price, combining regular (depromoted) price and any applicable short-term price cuts. Over the calibration period, these average prices are: Bestever \$1.76, Citrus Hill \$1.83, Tropicana Regular \$1.75, Tropicana Premium \$2.26, Minute Maid \$1.98, and Maplehurst \$1.33.

6.1. *Future Research*

Having highlighted the role of nonstationarity in the multinomial logit model, it is worth discussing some further uses of the conceptual and methodological issues presented in this paper. One promising application would be to examine the long-run effects of promotions on household purchasing behavior. Numerous researchers and practitioners have debated the short-term effects of consumer promotions on brand profitability and consumer preferences (Blattberg and Neslin 1989), but few have tried to empirically test these consequences beyond a period of roughly one year. The technique developed here provides a natural first step in this modeling process.

Of course, our measure is not without its limitations. The NSDM model captures nonstationarity as a process of repeated renewals from a stationary Dirichlet distribution. This implies that all individuals will have the same long-run choice probabilities.⁷ Thus, while our model may do well in tracking choice behavior over the short run, it may not be appropriate for long-range forecasting (especially when the estimated probability of renewal is relatively high).

There are also a number of possible extensions worth considering. We briefly discuss some of these potential directions:

More general forms of nonstationarity. As researchers gain further insight into the nature of nonstationarity, we might want to consider more general models to describe changes in tastes and choice behavior. For example, it is possible that the same consumer exhibits different kinds of nonstationarity. Under some circumstances we might expect to see abrupt changes in brand preferences (as in NSDM); in other cases we might expect to see more continuous, gradual changes (as in SMOOTH).

Heterogeneity in nonstationarity. An explicit assumption of the NSDM model is that the nonstationarity exhibited by each consumer is described by exactly the same renewal process. A less restrictive model would allow the nonstationarity parameter λ to vary across households. This might be done in one of two ways: either by using a semiparametric specification to approximate the underlying distribution of λ , or by making λ a function of household characteristics such as category usage frequency. While this extension is beyond the scope of the application presented here, it could aid the model's realism and performance in a very substantial yet parsimonious manner.⁸

Acknowledgements. The authors acknowledge the constructive suggestions of Don Morrison, the Area Editor and two anonymous reviewers. The authors also thank Doug Honnold of Information Resources, Inc., for providing the data used in this study.

⁷ Note that the same thing is true of the exponential smoothing model, which is in effect a random walk with no absorbing state: forecasting over the long-run, individual choice probabilities will regress to the aggregate market shares.

⁸ This paper was received on March 29, 1991, and has been with the authors 3 months for 3 revisions. Processed by Marcel Corstjens, Area Editor.

References

- Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press.
- Blattberg, Robert C. and Scott A. Neslin (1989), "Sales Promotion: The Long and the Short of It," *Marketing Letters*, 1 (December), 81-97.
- Chintagunta, Pradeep K., Dipak C. Jain and Naufel J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research*, 28 (November), 417-428.
- Ehrenberg, A. S. C. (1988), *Repeat Buying: Facts, Theory, and Applications*, London: Charles Griffin and Co.; New York: Oxford University Press.

- Fader, Peter S. (1993), "Integrating the Dirichlet-Multinomial and Multinomial Logit Models of Brand Choice," *Marketing Letters* (forthcoming).
- , James M. Lattin and John D. C. Little (1992), "Estimating Nonlinear Parameters in the Multinomial Logit Model," *Marketing Science*, 11 (Fall), 372–385.
- Guadagni, Peter M. and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203–238.
- Jeuland, Abel P., Frank M. Bass and Gordon P. Wright (1980), "A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes," *Operations Research*, 28 (March–April), 255–277.
- Jones, J. Morgan and Jane T. Landwehr (1988), "Removing Heterogeneity Bias from Logit Model Estimation," *Marketing Science*, 7 (Winter), 41–59.
- Krishnamurthi, Lakshman and S. P. Raj (1988), "A Model of Brand Choice and Purchase Quantity Price Sensitivities," *Marketing Science*, 7 (Winter), 1–20.
- Massy, William F., David B. Montgomery and Donald G. Morrison (1970), *Stochastic Models of Buying Behavior*, Cambridge, MA: MIT Press.
- Neslin, Scott A. and Robert W. Shoemaker (1989), "An Alternative Explanation for Lower Repeat Rates After Promotion Purchases," *Journal of Marketing Research*, 26 (May), 205–213.
- Ortmeyer, Gwen, James M. Lattin and David B. Montgomery (1991), "Individual Differences in Response to Consumer Promotions," *International Journal of Research in Marketing*, 8, 169–186.
- Quandt, Richard E. and Stephen M. Goldfeld (1987), "GQOPT4/I Version 4.00," Princeton, NJ: Department of Economics, Princeton University.
- Sabavala, Darius J. and Donald G. Morrison (1977), "A Model of TV Show Loyalty," *Journal of Advertising Research*, 17 (December), 35–43.
- and ——— (1981), "A Nonstationary Model of Binary Choice Applied to Media Exposure," *Management Science*, 27 (6), 637–657.
- Srinivasan, V. and R. Kesavan (1976), "An Alternate Interpretation of the Linear Learning Model of Brand Choice," *Journal of Consumer Research*, 3 (September), 76–83.
- Vanhonacker, Wilfried R. and Russell S. Winer (1990), "A Rational Random Behaviour Model of Choice," *Applied Stochastic Models and Data Analysis*, 6, 41–52.