



Estimating Nonlinear Parameters in the Multinomial Logit Model

Peter S. Fader; James M. Lattin; John D. C. Little

Marketing Science, Vol. 11, No. 4. (Autumn, 1992), pp. 372-385.

Stable URL:

<http://links.jstor.org/sici?sici=0732-2399%28199223%2911%3A4%3C372%3AENPITM%3E2.0.CO%3B2-G>

Marketing Science is currently published by INFORMS.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/informs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

ESTIMATING NONLINEAR PARAMETERS IN THE MULTINOMIAL LOGIT MODEL

PETER S. FADER, JAMES M. LATTIN AND JOHN D. C. LITTLE

University of Pennsylvania

Stanford University

Massachusetts Institute of Technology

Multinomial logit models, especially those calibrated on scanner data, often use explanatory variables that are nonlinear functions of the parameters to be estimated. A common example is the smoothing constant in an exponentially weighted brand loyalty variable. Such parameters cannot be estimated directly using commercially available logit packages. We provide a simple iterative method for estimating nonlinear parameters at the same time as the usual linear coefficients. The procedure uses standard multinomial logit software and, in experience to date, converges rapidly. We prove that, under suitable conditions, the resulting parameter values are maximum likelihood estimates and show how to calculate asymptotic standard errors from normal computer output. Three applications illustrate the method in practice.

(Choice Models; Estimation and Other Statistical Techniques)

1. Introduction

Over the past decade, many researchers working with household-level scanner data have turned to the multinomial logit (MNL) model as a powerful framework for studying choice decisions involving multiple alternatives. In addition to tracking brand choice behavior (e.g., Guadagni and Little 1983), researchers have also used MNL to analyze a broad array of related psychological and behavioral phenomena, such as reference price and promotion effects (Lattin and Bucklin 1989, Kalwani et al. 1990, Gurumurthy and Little 1989), purchase event feedback (Srinivasan and Kibarian 1989), variety seeking behavior (Lattin 1987), promotion responsiveness (Ortmeyer, Lattin and Montgomery 1991), and response to advertising (Tellis 1988).

In almost every case, these models involve relatively complicated constructs designed to capture aspects such as preference heterogeneity, price expectations, or advertising exposure. Often such variables involve prior household behavior (either purchase behavior or exposure to marketing mix activity such as price or advertising) in functional forms that are nonlinear in one or more parameters. For example, in order to account for the differences in brand preferences across households, Guadagni and Little (1983) create a variable, called brand loyalty, which is an exponentially weighted average of the past purchase history of the household. Fader (1992), using the Dirichlet-multinomial model, suggests another way of capturing cross-sectional heterogeneity. Abe (1991) develops an advertising stock variable that consists of an exponentially smoothed sum of past household exposures.

These models share a common problem: because the parameters of some variables are embedded nonlinearly, they cannot easily be estimated along with the other logit coefficients. As a result, few studies have utilized rigorous techniques to estimate values for these parameters. In some cases, researchers have simply chosen "reasonable" values based on prior studies and/or intuition. The costs of such simplifications are twofold. First, it is usually impossible to assess a sensible standard error for a preset parameter or determine its covariance with estimates of other coefficients. Second, obstacles to model calibration impede the development of new, more elaborate constructs. Indeed, it seems likely that many researchers would develop and use richer MNL models if the estimation process were more straightforward.

The purpose of this paper is to introduce a computationally efficient, iterative procedure that allows nonlinear parameters to be estimated using standard linear-in-parameters MNL software as a subroutine. We prove that, under appropriate conditions, the algorithm generates maximum likelihood estimates for the nonlinear parameters and we illustrate it in three practical applications. These provide examples of the procedure's convergence properties and show that inadequate estimation of a nonlinear parameter can affect other model coefficients. We also report on the experiences of others who have used this method.

The applicability and potential benefits of the technique will depend on the particular application. The basic requirements are (1) that the nonlinear functions of parameters to be estimated have derivatives that can be calculated reasonably efficiently for each data point, (2) that the functions be smooth in the nonlinear parameters near their maximum likelihood values, and (3) that the starting values for the parameters be sufficiently close to the global maximum of the likelihood function as to lie in a concave region containing the maximum. Since nonlinear functions cover many possibilities, one would assume that sometimes the requirements would not be met, but we have encountered very few difficulties across a considerable variety of applications.

2. Nonlinear Estimation Algorithm

The usual multinomial logit (MNL) model assumes that a linear combination of the attributes is linked to choice probabilities as follows:

$$p_j^h(t) = e^{v_j^h(t)} / \sum_k e^{v_k^h(t)}, \quad \text{where} \quad (1)$$

$p_j^h(t)$ = the probability that household h chooses brand j on purchase occasion t ,

$v_j^h(t) = \sum \beta_r x_{jr}^h(t)$

= the deterministic component of utility of brand j to household h at purchase occasion t ,

$x_{jr}^h(t)$ = r th explanatory variable for brand j and household h on purchase occasion t , $r = 1, \dots, R$,

β_r = coefficient to be estimated.

In applications to household scanner data, the $x_{jr}^h(t)$ generally include brand-specific intercept terms and marketing mix variables such as price and different types of promotions, and sometimes exposure to television advertising. In addition, variables may be added to capture other sources of variation across households and over time.

We first consider the case of a MNL model with any number of variables, of which one, $x_{jm}^h(t)$, is nonlinearly dependent on a single parameter α . Because α is imbedded within $x_{jm}^h(t)$, it cannot be estimated directly as an ordinary logit coefficient. For expositional clarity, we suppress the subscripts m , h and j , and make α explicit. The notation for $x_{jm}^h(t)$ becomes $x(t, \alpha)$.

First expand $x(t, \alpha)$ in a Taylor series around a starting value α_0 :

$$x(t, \alpha) = x(t, \alpha_0) + \left. \frac{dx(t, \alpha)}{d\alpha} \right|_{\alpha_0} (\alpha - \alpha_0) + \sum_{n=2}^{\infty} \left. \frac{d^n x(t, \alpha)}{d\alpha^n} \right|_{\alpha_0} \frac{(\alpha - \alpha_0)^n}{n!}. \quad (2)$$

If $x(t, \alpha)$ is smooth (e.g., its derivatives with respect to α are bounded) in an interval containing both α_0 and the maximum likelihood estimate (MLE) value of α , then the second and higher-order terms in (2) will approach 0 as α_0 approaches its MLE value. Letting $x'(t, \alpha) = dx(t, \alpha)/d\alpha$, we have as a current approximation for $x(t, \alpha)$,

$$x(t, \alpha) \simeq x(t, \alpha_0) + x'(t, \alpha_0)(\alpha - \alpha_0), \quad (3)$$

which becomes exact upon convergence of α_0 to α .

Letting β be the coefficient for $x(t, \alpha)$ in the MNL, the contribution of $x(t, \alpha)$ to utility is approximately

$$\beta x(t, \alpha) \simeq \beta x(t, \alpha_0) + \beta(\alpha - \alpha_0)x'(t, \alpha_0). \quad (4)$$

From (4), we see that we can better represent the contribution of $\beta x(t, \alpha)$ to utility by including $x'(t, \alpha_0)$ as well as $x(t, \alpha_0)$ among the variables in the MNL estimation. Denoting the resulting estimates by β' and β , we end up with a contribution to utility of

$$\beta x(t, \alpha) \simeq \beta x(t, \alpha_0) + \beta' x'(t, \alpha_0). \quad (5)$$

Comparing (4) and (5), we see that $\beta' \simeq \beta(\alpha - \alpha_0)$, or

$$\alpha \simeq \alpha_0 + \beta'/\beta. \quad (6)$$

Thus, we can use β' to obtain a new, better estimate of α . Substituting this for α_0 , we iterate until $(\alpha - \alpha_0)$ becomes as small as desired; i.e., until $\beta' \simeq 0$. Usually this requires only a few iterations.

Although each iteration makes use of the Taylor series approximation for $x(t, \alpha)$, by iteratively running the MNL linear estimation routine, these approximations converge to the exact value of the function. Thus we have our Nonlinear Estimation Algorithm (NEA):

- (1) Choose a starting value of α , say α_0 .
- (2) Calculate $x(t, \alpha)$ and $x'(t, \alpha)$ at α_0 for all observations t .
- (3) Include $x(t, \alpha_0)$ and $x'(t, \alpha_0)$ along with all the other variables in the logit model, and estimate coefficients in the usual manner.
- (4) Update α_0 using equation (6): $\alpha_0 \leftarrow (\alpha_0 + \beta'/\beta)$.
- (5) Return to step (2) and iterate until α_0 converges, i.e., until the coefficient of x' is indistinguishable from 0. Denote the final estimates by $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\beta}'$.
- (6) Calculate the standard error of $\hat{\alpha}$ from $SE(\hat{\alpha}) = SE(\hat{\beta}')/\hat{\beta}$.

Extension to multiple nonlinear parameters is straightforward. Suppose that one of the independent variables involves a J -dimensional vector of imbedded parameters: $x(t, \alpha)$, where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_J)$. The development now uses the linear terms of a multivariate Taylor series. The iterative estimation procedure still applies. The partial derivatives of $x(t, \alpha)$ with respect to the α_j are evaluated at some starting vector $\alpha_0 = (\alpha_{10}, \alpha_{20}, \dots, \alpha_{J0})$. All of the derivative variables are included in the logit model. The analog of (6) is

$$\alpha_j \simeq \alpha_{j0} + \beta'_j/\beta, \quad (7)$$

where β'_j represents the logit coefficient for the j th derivative variable, and β is the logit coefficient for $x(t, \alpha)$. The update in step (4) becomes $(\alpha_{j0} \leftarrow \alpha_{j0} + \beta'_j/\beta)$. If there are several variables containing nonlinear parameters, each variable and its parameters are treated in the same way as α .

Although the introduction of multiple nonlinear parameters requires nothing new theoretically or conceptually, the number of iterations needed for convergence tends to

increase. In a case where a model with a single α might only require 2–3 runs of the linear estimation program, our experience has shown that a set of 6 α_j 's takes about 5–7 iterations.

3. Satisfying Maximum Likelihood Conditions

We now show that the algorithm NEA, under conditions to be described, leads to maximum likelihood estimates for parameters that enter nonlinearly into variables of the MNL utility function. We restrict ourselves to nonlinear functions whose derivatives with respect to the nonlinear parameters have an upper and lower bound over the domain of interest for the parameters. This is not a serious limitation for practical applications. Because the range of possible functions is large, it is difficult to believe that the algorithm can be proven always to converge to a global maximum. We shall show instead that convergence, if it occurs, is at least to a local maximum. In practice, whenever we have investigated the matter, the maximum has turned out to be global.

The development builds on the fact that we are actually solving a series of familiar linear-in-parameters problems. The conditions required for the linear-in-parameters log likelihood function to have a unique global maximum (and therefore be guaranteed to produce maximum likelihood estimates) are described by McFadden (1974). The matrix of second derivatives of the log likelihood function must be nonsingular and negative definite. He shows under fairly weak conditions that, for the MNL, if a maximum exists, it will be unique and the global maximum.

Our argument proceeds as follows (formal proofs appear in the appendix). First, we show that the first-order conditions for a maximum in the nonlinear problem are satisfied when NEA has converged (Theorem 1). In doing so we assume that the coefficient of the variable containing the nonlinear parameter has an estimated value that is nonzero (i.e., $\hat{\beta} \neq 0$). In other words, the variable must have some influence on utility. If the coefficient were actually zero, the algorithm could not validly converge and, in fact, if its estimate is near zero, convergence may be problematic. At convergence, the derivative terms in the Taylor series disappear (i.e., $\hat{\beta}^t = 0$), showing that, although we use the Taylor series to set up the algorithm, the final result is exact.

Next we investigate the second-order conditions to determine whether or not the first-order conditions yield a maximum. We find that, at convergence, the Hessian of the nonlinear problem is negative definite (Lemma 1, Lemma 2, and Theorem 2), which implies that the estimates from NEA provide at least a local maximum.

Finally, formulas are developed for the standard errors of the estimated nonlinear parameters. Asymptotically, the variance-covariance matrix of the model parameter estimates is equal to the inverse of the Hessian for the nonlinear problem. Lemma 2 shows that we can obtain the inverse of the Hessian for the *nonlinear* problem directly from the inverse of the Hessian for the final *linear* problem. The resulting expression for the standard error appears in step (6) of NEA.

4. Applications

We now illustrate the implementation and performance of the proposed estimation technique with three quite different applications. The first finds the smoothing constant in a traditional exponentially smoothed loyalty variable. The second is a multivariate application: determining the parameters in a Dirichlet-multinomial model. Finally, we estimate a forgetting constant for an advertising response model.

The first two examples use scanner data on refrigerated orange juice provided by Information Resources, Inc. The database includes 3,079 scanned purchases made by 200 randomly chosen households in Marion, IN during the years 1983–84. We include the six most-purchased products over this period, accounting for 80% of all orange juice

purchases for our sample households. The six products are all in 64 ounce packages and include four national brands (Citrus Hill, Minute Maid, Tropicana Regular, and Tropicana Premium), one regional brand and one private label. The 1,490 purchases made in 1983 are used as an initialization period for the exponentially smoothed loyalty measure, while the remaining 1,589 purchases are used for model calibration.

4.1. Example 1: Exponential Smoothing Constant

The exponentially smoothed brand loyalty term may be expressed in the following form:

$$\text{LOY}_j^h(t) = \lambda \text{LOY}_j^h(t-1) + (1-\lambda)y_j^h(t-1), \quad \text{where} \quad (8)$$

$\text{LOY}_j^h(t)$ = loyalty of household h to brand j on purchase t ,

$y_j^h(t) = 1$ if household h buys brand j on purchase occasion t , 0 otherwise,

λ = smoothing parameter, $0 \leq \lambda \leq 1$.

This loyalty measure, introduced by Guadagni and Little (1983), accounts for differences in tastes across households and also tracks changes in tastes over time. Cross-validation studies with hold-out samples of households confirm the explanatory power of the variable. Many researchers have subsequently used this loyalty measure or some variant of it. Other ways of representing cross-sectional heterogeneity and purchase event feedback have been developed in what continues to be an active area of research: c.f., Chintagunta, Jain, and Vilcassim (1991); Lattin (1990); and Srinivasan and Kibarian (1989). Several of these alternative measures also involve nonlinear parameters and could therefore benefit from the proposed estimation method.

In their original application, Guadagni and Little (1983) used a fairly complex iterative scheme to get approximate values for λ . They began with an arbitrarily chosen value of $\lambda = 0.75$ to calculate coefficients for the full model. They next added dummy variables to capture the carryover effects for each of the ten most recent purchases by each household. Finally, they fit an exponential decay curve to these dummy variable coefficients to derive estimates of $\lambda_B = 0.875$ for brand loyalty and $\lambda_S = 0.812$ for package size loyalty.

Few researchers (an exception is Kannan and Wright 1991) have gone to this much trouble to determine values for exponential smoothing constants. While some have employed grid searches (e.g., Gurumurthy and Little 1989), most researchers simply choose a convenient value of λ (usually between 0.7 and 0.9) and make no attempt to refine it. Examples include Lattin (1987), Gupta (1988), Kalwani et al. (1990), Ortmeyer, Lattin and Montgomery (1991), and Papatla and Krishnamurthi (1992).

In contrast to these ad hoc approaches, we now show how to obtain exact maximum likelihood estimates for λ using the algorithm described above.

Model specification. We first find $\text{DLOY}_j^h(t)$, the first derivative of $\text{LOY}_j^h(t)$ with respect to λ . Equation (8) can be re-expressed as

$$\text{LOY}_j^h(t) = (1-\lambda) \sum_{s=0}^{t-1} \lambda^s y_j^h(t-s-1). \quad (9)$$

Differentiating with respect to λ and collecting terms yields

$$\text{DLOY}_j^h(t) = \lambda \text{DLOY}_j^h(t-1) + \text{LOY}_j^h(t-1) - y_j^h(t-1). \quad (10)$$

Equation (10) is similar to (8). Given λ and initial conditions, both $\text{DLOY}_j^h(t)$ and $\text{LOY}_j^h(t)$ are easily computed recursively from the data and, in fact, can be done at the same time.

For initial conditions, since we lack information prior to each household's first purchase occasion, we have assumed equal loyalties: for each brand we set $\text{LOY}_j^h(1)$ equal to

$1/J$, where J is the number of brands. It follows that $DLOY_j^h(1) = 0$. For all subsequent purchase occasions, equations (8) and (10) apply. This initialization differs slightly from Guadagni and Little (1983), who set $LOY_j^h(t) = \lambda$ for the brand actually chosen on the first purchase occasion and $(1 - \lambda)/(J - 1)$ for all others.

The full MNL model in our example includes LOY, its derivative, DLOY, and variables for regular (depromoted) price, short-term price cuts, and a 0-1 variable for any feature activity. The linear coefficients are estimated by maximum likelihood using a standard FORTRAN Newton-Raphson algorithm for MNL adapted from Ben Akiva (1973).

Results. Table 1 demonstrates the efficacy of the iterative procedure for estimating λ in this example. The first model (iteration 0) illustrates the typical usage in the literature of a brand loyalty with a fixed smoothing constant (here $\lambda = 0.75$) and no updating. The fit is good and the coefficients significant.

In iteration 1, we start over using our procedure with $\lambda_0 = 0.75$, now including the DLOY term to allow updating. Equation (6) gives an increase in λ_0 of $0.22/3.72$, or 0.059 . Although this is less than 10% of λ_0 , the model shows a substantial improvement in fit ($\chi^2 = 10.70$, $p < 0.001$) over the base case. The logit coefficients change relatively little, with the exception of the coefficient for LOY.

Over the next few iterations, the log likelihood and estimated value of λ continue to show small positive increases. By the fifth iteration, both have converged within three decimal places of their ultimate values, and model coefficients remain virtually unchanged. The results are quite typical of what the authors have observed across a variety of data sets.

Robustness of convergence. The rapid convergence in this case might, of course, be due to a good initial choice of λ_0 . To examine this issue, we have rerun the model with beginning values of λ_0 ranging from 0.05 to 0.95 in increments of 0.1. In the worst case ($\lambda_0 = 0.05$), the procedure requires six iterations to reach an estimate of λ within 0.001

TABLE 1
Estimation Results for Brand Loyalty Smoothing Constant

	Iteration:					
	0	1	2	3	4	5
Initial smoothing constant (λ_0)	0.750	0.750	0.809	0.825	0.829	0.831
Logit coefficients: ^a						
Brand loyalty (LOY)	3.46	3.72	3.88	3.92	3.92	3.92
Brand loyalty derivative (DLOY)	—	0.22	0.06	0.02	0.005	0.001
(std. error for DLOY)	—	(0.065)	(0.058)	(0.055)	(0.055)	(0.054)
Regular price	-2.44	-2.47	-2.48	-2.48	-2.48	-2.48
Price cut	2.34	2.34	2.34	2.34	2.34	2.34
Feature	0.61	0.61	0.61	0.61	0.61	0.61
Brand-specific constants:						
Citrus Hill	1.08	1.13	1.16	1.16	1.16	1.16
Minute Maid	1.08	1.11	1.13	1.13	1.13	1.13
Tropicana Premium	0.84	0.87	0.88	0.88	0.88	0.88
Tropicana Regular	0.36	0.39	0.40	0.40	0.40	0.40
Regional brand	0.21	0.22	0.22	0.22	0.22	0.22
Private label brand ^b	0.00	0.00	0.00	0.00	0.00	0.00
Updated smoothing constant (λ)	—	0.809	0.825	0.829	0.831	0.831
(std. error for λ)	—	(0.017)	(0.015)	(0.014)	(0.014)	(0.014)
Log likelihood	-1446.34	-1440.99	-1437.85	-1437.63	-1437.61	-1437.61

^a With the exception of the coefficients for DLOY, all coefficients are significant at $p = 0.10$.

^b Brand-specific constant for private label brand constrained to 0.00.

of the final value; in the best ($\lambda_0 = 0.85$) only two. Although this convergence is very good and typifies our usual experience, there have occasionally been cases which have required trial-and-error search for a suitable starting value of λ_0 .

Global optimality. To assess the global optimality of our results, we have performed a grid search. Without using the Taylor series technique, we let λ vary from 0 to 1 with increments of 0.01 and graph the resulting log likelihoods. The results, shown in Figure 1, clearly show the presence of a single peak near 0.83. We then narrowed the range from 0.75 to 0.90 and conducted another grid search using increments of 0.001. These results are also graphed in Figure 1.

Besides confirming the optimality of the point of convergence for the technique, this analysis offers additional insight. The width of the plateau around the optimum point is surprisingly narrow. We compared the log likelihoods in Figure 1 to the optimal value from Table 1 (-1437.61) using a χ^2 test with 1 d.f. (to account for the added derivative term). Only those values of λ in the range $(0.781, 0.871)$ pass this test at the 95% level of significance. This stands in contrast to the reports by some researchers, for example, Gupta (1988), who tried smoothing constants of 0.7, 0.8, and 0.9 and found that results were insensitive to these changes. The tightness of the range in Figure 1 is an important argument for considering an accurate estimation technique.

Effects on other model coefficients. Although the primary purpose of our method is simply to obtain the right answer (i.e., put the nonlinear parameters on the same maximum likelihood footing as the linear ones), it is worth noting that this technique has further value. In almost all cases the value of λ will have some effect on the other coefficients and in special cases the effect can be quite substantial.

First we note in Table 1 that six of the nine estimated coefficients show increases in absolute value as the iterations converge. This is a predictable consequence of explaining more variance with the multinomial logit and is obviously desirable.

Next, any variable that is partially collinear with LOY will tend to be sensitive to λ . We illustrate such a situation with a lagged promotional purchase dummy variable, LAGPROM. Guadagni and Little (1983) included such a variable in their model. Table 2 shows estimation results for a pair of MNL models that include LAGPROM with λ fixed at 0.7 and 0.9. The coefficients of LAGPROM are significant (at $p < 0.10$) in both cases but have different signs! The actual MLE value of λ (for the model including LAGPROM) is 0.829, at which point the coefficient for LAGPROM is nonsignificant

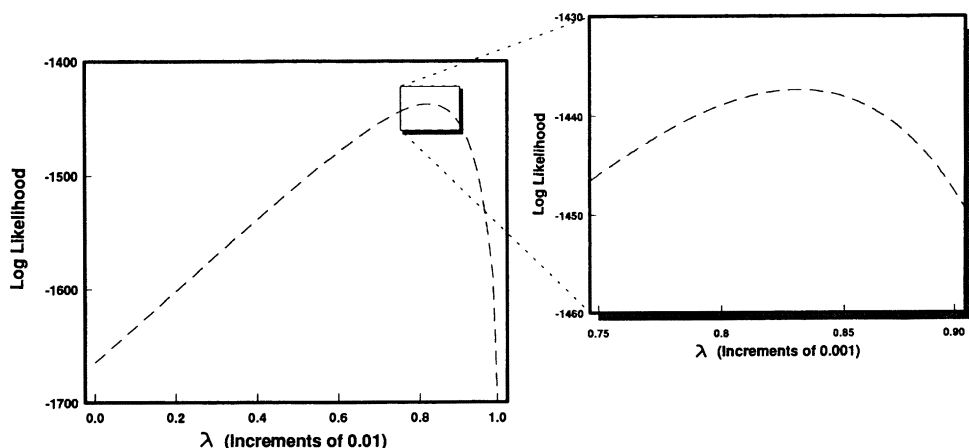


FIGURE 1. Model Fits for Varying λ .

(−0.040 with *t*-statistic 0.33). Thus, if collinearity is present, arbitrarily presetting the nonlinear parameters is particularly dangerous.

4.2. Example 2: Dirichlet-Multinomial Model

The Dirichlet-multinomial (DM) model is a well-accepted method of accounting for cross-sectional heterogeneity in repeat-purchase situations, especially when explanatory (marketing mix) variables are unavailable. We assume that our priors about the probabilities governing the choice behavior of any household are described by a Dirichlet distribution, the multivariate analog of the Beta distribution. Using the observed choice behavior of the household before time *t* (i.e., the outcomes of a multinomial choice process), we can update our priors to obtain a posterior estimate of the probability of choosing brand *i*, given by

$$\frac{\alpha_j + n_j^h(t - 1)}{\sum_k (\alpha_k + n_k^h(t - 1))},$$

(11)

where $n_j^h(t - 1)$ is the number of choices of brand *j* made by household *h* through time *t* − 1.

The brand-specific parameters, $\alpha_1, \alpha_2, \dots, \alpha_J$, capture the relative popularity of each brand, and can also be used to assess the overall level of preference heterogeneity in the market. For further background and applications of the DM model, see Goodhardt, Ehrenberg and Chatfield (1984).

Model specification. Fader (1992) shows that the DM model can be calibrated within the MNL framework by including the following single variable in the logit utility function with coefficient constrained to 1.0:

$$DM_j^h(t) = \ln (\alpha_j + n_j^h(t - 1)).$$

(12)

Assuming no marketing mix effects, substituting (12) into the basic MNL model in (1) yields exactly the expected purchase probability in (11).

TABLE 2

Estimation Results with Lagged Promotional Purchase Variable

	$\lambda = 0.70$	$\lambda = 0.90$
Logit coefficients: ^a		
Brand loyalty (LOY)	3.41	4.53
Regular price	−2.43	−2.52
Price cut	2.34	2.34
Feature	0.62	0.59
Lagged promotional purchase (LAGPROM)	−0.32	0.19
(<i>t</i> -statistic for LAGPROM)	(−2.73)	(1.67)
Brand-specific constants:		
Citrus Hill	1.03	1.27
Minute Maid	1.05	1.22
Tropicana Premium	0.79	0.95
Tropicana Regular	0.35	0.44
Regional brand	0.21	0.22
Private label brand ^b	0.00	0.00
Log likelihood	−1452.96	−1449.34

^a All coefficients are significant at *p* = 0.10.

^b Brand-specific constant for private label brand constrained to 0.00.

Estimation is only required for the J Dirichlet parameters (α_j 's). The multivariate version of our nonlinear estimation algorithm will find them. First calculate the partial derivative of $DM_j^h(t)$ with respect to each of the α 's:

$$DDM_{ij}^h(t) = \frac{\partial DM_j^h(t)}{\partial \alpha_i} = \begin{cases} 1/(\alpha_j + n_j^h(t - 1)) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{13}$$

Each of these derivatives is evaluated at a vector of starting values α_0 (market shares are often a good choice) and the following logit model is estimated:

$$v_j^h(t) = DM_j^h(t) + \beta_1 DDM_{1j}^h(t) + \beta_2 DDM_{2j}^h(t) + \cdots + \beta_J DDM_{Jj}^h(t). \tag{14}$$

The updated DM parameter estimates are calculated according to equation (7), and the logit model in (14) is iteratively estimated until convergence.

Results. Estimation results are shown in Table 3. By the sixth iterative run of the MNL model, all of the relevant statistics are accurate to three significant digits. More remarkable is the quality of the model after only one iteration. Although the α_j 's seem far from convergence, the model fit is not significantly different from its optimal value. Nevertheless, it is generally worthwhile to perform 2–3 more iterations to ensure that the α_j 's are within 5% of their final values.

The DM parameter estimates stand up quite well to the types of tests performed in the previous example. To examine the robustness of the estimation procedure, we have tried starting values ranging from 0.001 to as high as 10.0 for each brand, and in all cases the α 's quickly converged to their correct maximum likelihood estimates.

Many other interesting MNL models with nonlinear parameters grow out of this example. The pure DM model shown here does not contain any marketing mix variables in the logit formulation. However, nothing prevents the model-builder from introducing them. If price and promotion variables are added, then DM_j^h can be treated as a loyalty variable in place of the exponentially smoothed measure. Fader (1991) motivates such a model and compares it to the two models (pure DM and smoothed loyalty) shown here. Hardie, Johnson and Fader (1991) use this hybrid DM-MNL combination to operationalize a new model by Tversky and Kahneman (1991) featuring loss aversion and reference-dependent choice. Fader and Lattin (1992) go further in employing a stand-alone stochastic choice model as a loyalty measure within the MNL model, using a multivariate extension of the beta-binomial-geometric model (Sabavala and Morrison 1981). Each of these models relies upon the estimation procedure outlined here.

TABLE 3
Estimation Results for Dirichlet-Multinomial Model

Brand	0*	Iteration:					
		1	2	3	4	5	6
Citrus Hill	0.288	0.432	0.489	0.501	0.505	0.506	0.506
Minute Maid	0.236	0.356	0.397	0.404	0.407	0.407	0.408
Regional brand	0.151	0.263	0.315	0.322	0.324	0.324	0.324
Tropicana Regular	0.146	0.271	0.339	0.350	0.352	0.353	0.353
Private label	0.137	0.309	0.475	0.556	0.579	0.583	0.584
Tropicana Premium	0.042	0.068	0.080	0.082	0.083	0.083	0.083
Log likelihood	-1751.58	-1716.83	-1715.34	-1714.68	-1714.58	-1714.57	-1714.57

* Initial estimates are market shares.

4.3. Example 3: An Advertising Response Model

Scanner panel datasets that include TV advertising exposures at the household level offer excellent opportunities for modeling market response to advertising. However, many hypothesized advertising phenomena, such as response thresholds, diminishing returns, forgetting, and copy wear-out, are likely to require functions in which the parameters enter nonlinearly. As an example, Abe (1991) assesses consumer response to advertising by constructing a cumulative measure of exposure that he calls “advertising stock.” This is included as a variable in a logit choice model. He calibrates the MNL using a scanner panel database for cranberry drinks. The advertising data consist of instants (to the minute) at which a panel household had a TV set turned on and a product commercial was shown. His model supposes that such an advertising exposure has its greatest impact on consumer utility immediately and then has a declining influence in a forgetting process represented by an exponential decay. The stock of past advertising that affects a customer’s consumer’s utility is taken to be the sum of partially retained past exposures.

Thus, if t_0 is the time of the last purchase occasion and $\{t_i\}$ the instants of advertising exposure since t_0 , the advertising stock of household h for product j at time t is expressed as

$$\text{ADSTOCK}_j^h(t) = \gamma^{t-t_0} \text{ADSTOCK}_j^h(t_0) + \sum_i \gamma^{t-t_i}. \quad (15)$$

Time can be measured, for example, in days. If a sudden burst of advertising creates multiple exposures in a household, adstock will rise quickly but, in the absence of further exposures, will decay towards zero.

The adstock parameter (γ) captures the retention rate of advertising. One may use it to determine the “half-life” for an ad exposure, i.e., the time until a consumer retains only half the impact from the original exposure. Estimates for γ tend to be near or above 0.9 when time is measured in days. For example, a γ of 0.9 would imply a half-life of 6.6 days for a single exposure. The coefficient of adstock, in conjunction with γ , determines the effectiveness of advertising in influencing consumer utility and therefore brand choice. A calibrated logit model that includes advertising offers the possibility of building advertising planning models based on micro level response phenomena.

Notice that, although advertising stock represents an exponential smoothing of past exposures, it has a more complicated analytic form than a Guadagni-Little loyalty variable because it is based on absolute time rather than number of purchase occasions. In addition, there are many more advertising exposures than purchases. The variable therefore presents a new challenge for the estimation technology.

Abe (1991) calibrates his model on an IRI scanner panel database consisting of purchases of 15 brand-sizes of cranberry drinks by 194 households over 104 weeks. A total of 989 purchases were made in this period. An additional 371 purchases in a 52-week pre-period were used for initialization. The advertising exposures were measured in 52 of the weeks with a total of 7,474 exposures recorded. In addition to the advertising stock variable, Abe’s model includes an exponential smoothing model of brand loyalty, in-store display (coded as a 0, 1 dummy), newspaper features (coded as 0, 1, 2, 3 based on the size of the ad), price per ounce, and a set of alternative specific constants. Both the loyalty and ad stock variables contain nonlinear parameters, which are simultaneously estimated.

Table 4 shows the results. The method converges to give the nonlinear parameters with three decimal place accuracy in five iterations. As may be seen, the value of γ is 0.909, implying a rather rapid decay of individual exposures. However, the coefficient of adstock shows a significant ($p < 0.10$) overall effect of advertising. Abe (1991) goes on to build these micro results into a macro decision model.

TABLE 4
Estimation Results for Advertising Response Model

	Iteration:					Final Model
	1	2	3	4	5	
Advertising retention (γ)	0.9500	0.9310	0.9102	0.9085	0.9094	0.9090
Loyalty smoothing constant (λ)	0.8000	0.7813	0.7781	0.7778	0.7777	0.7777
Logit coefficients: ^a						
Adstock	0.100	0.120	0.117	0.116	0.117	0.117
Adstock derivative	-0.002	-0.003	-0.0002	0.0001	-0.0001	
(std. error for derivative)	(0.003)	(-0.003)	(0.010)	(0.011)	(0.010)	
Brandsize loyalty	5.318	5.301	5.299	5.299	5.299	5.299
Brandsize loyalty derivative	-0.100	-0.017	-0.002	-0.0002	-0.0001	
(std. error for derviative)	(0.084)	(0.087)	(0.088)	(0.088)	(0.088)	
Price	-0.980	-0.979	-0.978	-0.978	-0.978	-0.978
Display	1.067	1.068	1.068	1.068	1.068	1.068
Feature	0.377	0.376	0.375	0.375	0.375	0.375
(Brandsize-specific constants omitted)						
Log likelihood	-1404.27	-1404.07	-1404.09	-1404.08	-1404.08	-1404.08

^a All coefficients in final model are significant at $p < 0.10$.

5. Discussion

Two of this paper's authors introduced the nonlinear estimation procedure to analysts at Information Resources in 1987, who have used the technique hundreds of times since then for a variety of nonlinear parameters. Several product categories have been examined multiple times (i.e., with new datasets) and have shown strong consistency in their nonlinear parameter estimates. When a particular product category is analyzed on a repeated basis, its prior parameter estimates are reused as starting values for the new model. These refined starting values almost never require more than 1 iteration before the derivative term's t -statistic goes below 2.0.

Besides the brand loyalty measure discussed here, several other exponentially smoothed variables have been frequently helpful in scanner databases. Guadagni and Little (1983) included *size loyalty* (i.e., observed tendencies to choose a particular package size). *Manufacturer loyalty* is valuable in cases where one manufacturer has two or more different brands in the same category. In addition, some categories require one or more *type loyalties*, which cover a variety of other product attributes. A complete analysis of the toothpaste market, for example, might include type loyalty terms for product form (paste/gel), package shape (tube/pump), flavor (regular/mint/other), and strength (regular/tartar control).

Parameter estimation for such MNL models would be quite difficult without the nonlinear estimation procedure used here. The alternative techniques mentioned earlier would be extremely cumbersome and time-consuming. Even the most efficiently implemented grid search (using finer grids in successive stages of the search) would require hundreds of separate model calibrations. Alternatively, researchers could abandon the commercially available logit packages, and program their own likelihood functions using full-information maximum likelihood. This option has become more feasible lately with the availability of generalized, nonlinear optimization software (such as GAUSS or MATLAB) for personal computers. What remains to be seen is how the convergence properties and the efficiency of full-information maximum likelihood compares to the iterative estimation procedure presented here.

This paper has presented an easily implemented technique for estimating parameters that enter nonlinearly into the variables of multinomial logit utility functions. The procedure has a wide range of applications, three of which were discussed here in depth. Various tests have showed the technique to converge quickly and accurately. Further, we have found that, under some conditions, the use of the technique not only improves

overall model fit but also affects the estimated coefficients for other variables. We expect that other researchers may wish to consider using this methodology, both in the traditional multinomial logit setting and in new modeling approaches.¹

Acknowledgements. The authors thank Doug Honnold of Information Resources Inc. for his suggestions and help in making the data available for this study, and thank Makoto Abe for providing the results for one of the examples used here.

¹ This paper was received December 26, 1990, and has been with the authors 5 months for 2 revisions. Processed by Dominique Hanssens.

Appendix

For clarity and simplicity, consider one nonlinear parameter α and assume it appears in the m th explanatory variable $x_{jm}(t, \alpha)$ whose linear coefficient is β_m . Extension to more nonlinear parameters is straightforward. Let $x'_{jm}(t, \alpha) = dx_{jm}/d\alpha$ with linear coefficient β'_m , and let L denote the log likelihood function, where

$$L = \log \prod_{ij} [p_j(t)]^{y_{ij}(t)} = \sum_{ij} y_{ij}(t) \log p_j(t).$$

Introducing notation that makes explicit the dependence on the observations t and the utilities $\{v_j(t)\}$, let

$$L(t) \equiv L(v_1(t), \dots, v_J(t)) \equiv \sum_j y_j(t) \log p_j(t)$$

so that $L = \sum_t L(t)$.

We shall speak of the *nonlinear problem* as the task of finding values for $\theta^{(n)} \equiv (\beta_1, \dots, \beta_R, \alpha)$ to maximize L and, correspondingly, of the *final linear problem* as finding values of $\theta^{(l)} \equiv (\beta_1, \dots, \beta_R, \beta'_m)$ to maximize L in the last linear-in-parameters iteration of NEA. Parameter values at convergence will be indicated by a hat ($\hat{\cdot}$). When we discuss second-order conditions, we let H be the Hessian matrix of second partial derivatives of L with respect to the elements of θ , and \hat{H} be H evaluated at $\hat{\theta}$.

THEOREM 1. *If NEA converges and $\hat{\beta}_m \neq 0$, the resulting estimates $\hat{\theta}^{(n)} = (\hat{\beta}_1, \dots, \hat{\beta}_R, \hat{\alpha})$ satisfy the first-order conditions (FOC) for the nonlinear problem.*

PROOF. The first-order conditions for the β 's are

$$\frac{\partial L}{\partial \beta_s} = \sum_t \frac{\partial L(t)}{\partial \beta_s} = \sum_t \sum_j \frac{\partial L(t)}{\partial v_j} \frac{\partial v_j}{\partial \beta_s} = 0$$

or

$$\sum_{ij} \frac{\partial L(t)}{\partial v_j} x_{js}(t) = 0 \quad \forall s. \quad (\text{FOC } 1)$$

To these we must add the condition for α :

$$\frac{\partial L}{\partial \alpha} = \sum_t \sum_j \frac{\partial L(t)}{\partial v_j} \left[\beta_m \frac{\partial x_{jm}(t, \alpha)}{\partial \alpha} \right] = 0$$

or, since we assume that $\hat{\beta}_m \neq 0$,

$$\sum_{ij} \frac{\partial L(t)}{\partial v_j} x'_{jm}(t, \alpha) = 0. \quad (\text{FOC } 2)$$

However, NEA has found values $\hat{\theta}^{(l)} = (\hat{\beta}_1, \dots, \hat{\beta}_R, \hat{\beta}'_m)$ with $\hat{\beta}'_m = 0$ that satisfy the first-order conditions of the final linear problem. These include all of (FOC 1) and, in addition, a first-order condition for β'_m , namely,

$$\frac{\partial L}{\partial \beta'_m} = \sum_{ij} \frac{\partial L(t)}{\partial v_j} x'_{jm}(t, \alpha_0) = 0.$$

This is just (FOC 2) and so the parameter estimates satisfy all the first-order conditions (FOC 1, 2) of the nonlinear problem. \square

LEMMA 1. *If NEA converges, then for the resulting parameter estimates $\hat{\theta}^{(n)} = (\hat{\beta}_1, \dots, \hat{\beta}_R, \hat{\alpha})$ of the nonlinear problem,*

$$\sum_{ij} \frac{\partial L(t)}{\partial v_j} x''_{jm}(t, \alpha) = 0$$

where $x''_{jm} = d^2 x_{jm}/d\alpha^2$.

PROOF. Nowhere previously in the development have we introduced x'' and we are sorry to do so now, but it turns up in the diagonal of H and we must be prepared to get rid of it. The proof will be by construction. Suppose we had introduced x'' by adding another term in the Taylor's expansion for $x(t, \alpha)$. Equation (4) would have been:

$$\beta x(t, \alpha) \simeq \beta x(t, \alpha_0) + \beta(\alpha - \alpha_0)x'(t, \alpha_0) + \frac{1}{2}\beta(\alpha - \alpha_0)^2 x''(t, \alpha_0)$$

and (5):

$$\beta x(t, \alpha) \simeq \beta x(t, \alpha_0) + \beta' x'(t, \alpha_0) + \beta'' x''(t, \alpha_0).$$

In running the MNL estimation subroutine within NEA, we would have included x'' as a variable with coefficient β''_m . This gives the match-ups

$$\beta' = \beta(\alpha - \alpha_0) \quad \text{and} \quad \beta'' = \frac{1}{2}\beta(\alpha - \alpha_0)^2 = \frac{1}{2}\beta'(\alpha - \alpha_0).$$

On convergence, we would have obtained $\alpha = \alpha_0$, and $\beta' = 0$ as before and, in addition, $\beta'' = 0$. At each step along the way and at the end, the MNL program would have dutifully calculated β''_m , even though its contribution to utility is ultimately zero. The MNL program would have calculated β''_m in the final linear problem by enforcing a first-order condition:

$$\frac{\partial L}{\partial \beta''_m} = \sum_{ij} \frac{\partial L(t)}{\partial v_j} x''_{jm}(t, \alpha) = 0.$$

This is the desired result and proves the lemma. (For completeness, we note a nonproblem: Conceivably the inclusion of x'' as a variable could make the set of explanatory variables linearly dependent, thereby causing NEA not to converge when it otherwise would. However, in such a case, x'' can be expressed as a linear combination of the original variables. Substituting this for x'' and applying the first-order conditions for the original variables establishes the lemma for this special case.) \square

To discuss second-order conditions, i.e., whether or not the solution to the first-order conditions yields a maximum, we must examine second derivatives. Let $H^{(l)}$ be the Hessian for the final linear problem and $H^{(n)}$ be that for the nonlinear problem. Further, let $\hat{H}^{(l)}$ and $\hat{H}^{(n)}$ be their numerical values upon convergence of NEA.

LEMMA 2. $\hat{H}^{(n)}$ can be obtained from $\hat{H}^{(l)}$ by multiplying its last row and last column by $\hat{\beta}_m$.

PROOF. From the definition of a Hessian, the elements of $H^{(l)}$ are

$$\begin{aligned} h_{rs}^{(l)} &= \frac{\partial^2 L}{\partial \beta_r \partial \beta_s} = \sum_{ijk} x_{jr}(t) \frac{\partial^2 L(t)}{\partial v_j \partial v_k} x_{ks}(t), \quad r, s = 1, \dots, R, \\ h_{r, R+1}^{(l)} &= \frac{\partial^2 L}{\partial \beta_r \partial \beta'_m} = \sum_{ijk} x_{jr}(t) \frac{\partial^2 L(t)}{\partial v_j \partial v_k} x'_{km}(t, \alpha), \quad r = 1, \dots, R, \\ h_{R+1, s}^{(l)} &= \frac{\partial^2 L}{\partial \beta'_m \partial \beta_s} = h_{s, R+1}^{(l)}, \quad s = 1, \dots, R, \\ h_{R+1, R+1}^{(l)} &= \frac{\partial^2 L}{\partial \beta'^2_m} = \sum_{ijk} x'_{jm}(t, \alpha) \frac{\partial^2 L(t)}{\partial v_j \partial v_k} x'_{km}(t, \alpha). \end{aligned}$$

Correspondingly, the elements of $\hat{H}^{(n)}$ are

$$\begin{aligned} h_{rs}^{(n)} &= \frac{\partial^2 L}{\partial \beta_r \partial \beta_s} = h_{rs}^{(l)}, \quad r, s = 1, \dots, R, \\ h_{r, R+1}^{(n)} &= \frac{\partial^2 L}{\partial \beta_r \partial \alpha} = \sum_{ijk} x_{jr}(t) \frac{\partial^2 L(t)}{\partial v_j \partial v_k} \beta_m x'_{km}(t, \alpha) = \beta_m h_{r, R+1}^{(l)}, \quad r = 1, \dots, R, \\ h_{R+1, s}^{(n)} &= \frac{\partial^2 L}{\partial \alpha \partial \beta_s} = h_{s, R+1}^{(n)} = \beta_m h_{s, R+1}^{(l)}, \quad s = 1, \dots, R, \\ h_{R+1, R+1}^{(n)} &= \frac{\partial^2 L}{\partial \alpha^2} = \sum_{ijk} \beta_m x'_{jm}(t, \alpha) \frac{\partial^2 L(t)}{\partial v_j \partial v_k} \beta_m x'_{km}(t, \alpha) + A \\ &= \beta_m^2 h_{R+1, R+1}^{(l)} + A \quad \text{where} \\ A &= \sum_{ij} \frac{\partial L(t)}{\partial v_j} \beta_m x''_{jm}(t, \alpha). \end{aligned}$$

Examination of these relations shows that Lemma 2 holds if $A = 0$. However, Lemma 1 tells us that $A = 0$ on convergence, i.e., when $\theta^{(l)} = \hat{\theta}^{(l)}$, $\theta^{(n)} = \hat{\theta}^{(n)}$, $H^{(l)} = \hat{H}^{(l)}$ and $H^{(n)} = \hat{H}^{(n)}$. \square

THEOREM 2. *If NEA converges to a unique global maximum for the linear problem, then $\theta^{(n)} = (\hat{\beta}_1, \dots, \hat{\beta}_R, \hat{\alpha})$ yields a maximum for the nonlinear problem.*

PROOF. Since $\hat{\theta}^{(l)} = (\hat{\beta}_1, \dots, \hat{\beta}_R, \hat{\beta}'_m)$ determines a unique global maximum, expansion of L about this point will give a negative definite quadratic form for $\hat{H}^{(l)}$. But $\hat{H}^{(n)}$ differs from $\hat{H}^{(l)}$ only by having its last row and column multiplied by $\hat{\beta}_m \neq 0$. It is easily shown with linear algebra that $\hat{H}^{(n)}$ will then also be negative definite.

From this it follows that $\hat{\theta}^{(n)}$ is a maximum, at least locally. As we move away from this point, higher-order terms in the Taylor's expansion of L come into play and involve higher derivatives of $x(t, \alpha)$. We cannot unequivocally assert that these will not create another higher maximum but our restriction that the derivatives be bounded above and below means that, for $\theta^{(n)}$ sufficiently close to $\hat{\theta}^{(n)}$, $L(\theta^{(n)}) < L(\hat{\theta}^{(n)})$. \square

References

- Abe, Makoto (1991), "A Marketing Mix Model Developed from Single Source Data: A Semiparametric Approach," unpublished Ph.D. thesis, Cambridge, MA: Massachusetts Institute of Technology.
- Ben Akiva, Moshe (1973), "Program for Maximum Likelihood Estimation of the Multinomial Logit Model," Transportation Systems Division, Department of Civil Engineering, Cambridge, MA: Massachusetts Institute of Technology.
- 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.
- Fader, Peter S. (1992), "Integrating the Dirichlet-Multinomial and Multinomial Logit Models of Brand Choice," Department of Marketing Working Paper #91-013, Wharton School, University of Pennsylvania, Philadelphia, PA.
- and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, forthcoming.
- Goodhardt, G. J., A. S. C. Ehrenberg and C. Chatfield (1984), "The Dirichlet: A Comprehensive Model of Buying Behavior," *Journal of the Royal Statistical Society*, 147, 621–655.
- 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.
- Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342–355.
- Gurumurthy, K. and John D. C. Little (1989), "A Price Response Model Developed from Perceptual Theories," Sloan School Working Paper #3038-89, MIT Sloan School of Management, Cambridge, MA.
- Hardie, Bruce G. S., Eric J. Johnson and Peter S. Fader (1991), "Modeling Loss Aversion and Reference Brand Effects on Brand Choice" Department of Marketing Working Paper #91-025, Wharton School, University of Pennsylvania, Philadelphia, PA.
- Kalwani, Manohar U., Chi Kin Yim, Heikki J. Rinne and Yoshi Sugita (1990), "A Price Expectations Model of Customer Brand Choice," *Journal of Marketing Research*, 27 (August), 251–262.
- Kannan, P. K. and Gordon P. Wright (1991), "Modeling and Testing Structured Markets: A Nested Logit Approach," *Marketing Science*, 10 (Winter), 58–82.
- Lattin, James M. (1987), "A Model of Balanced Choice Behavior," *Marketing Science*, 6 (Winter), 48–65.
- (1990), "Measuring Preference from Scanner Panel Data: Filtering Out the Effects of Price and Promotion," unpublished working paper, Graduate School of Business, Stanford University, Stanford, CA.
- and Randolph E. Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299–310.
- McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, *Frontiers in Econometrics*, New York: Academic Press.
- Ortmeyer, Gwen, James M. Lattin and David B. Montgomery (1991), "Individual Differences in Response to Consumer Promotions," *International Journal of Research in Marketing*, 8 (September), 169–186.
- Papatla, Puroshottam and Lakshman Krishnamurthi (1992), "A Probit Model of Choice Dynamics," *Marketing Science*, 11 (Spring), 189–206.
- Sabavala, Darius J. and Donald G. Morrison (1981), "A Nonstationary Model of Binary Choice Applied to Media Exposure," *Management Science*, 27, 6, 637–657.
- Srinivasan, V. and Thomas Kibarian (1989), "Purchase Event Feedback: Fact or Fiction?," paper presented at the Marketing Science Conference, March 15–18, Fuqua School of Business, Duke University, Durham, NC.
- Tellis, Gerard J. (1988), "Advertising Exposure, Loyalty, and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25 (May), 134–144.
- Tversky and Kahneman (1991), "Loss Aversion and Riskless Choice: A Reference Dependent Model," *Quarterly Journal of Economics*, 106 (November), 1039–1061.