Marketing Letters 13:4, 373–388, 2002

A Guadagni–Little Likelihood Can Have Multiple Maxima

ERIC T. ANDERSON eric.anderson@gsb.uchicago.edu
University of Chicago, GSB, 1101 East 58th Street, Chicago, IL 60637

Accepted January 2002

Abstract

Despite many advances in marketing models, the Guadagni–Little (1983) model is still in widespread use by both practitioners and academics. For many new marketing models, the Guadagni–Little model serves as a benchmark. The key variable that allows the Guadagni–Little model to accurately fit data is the loyalty variable, which is an exponential smoothing of past purchases. In this paper, I show that inclusion of this variable in the logit model may result in a likelihood function that can have multiple maxima. I am able to demonstrate this using simulated data and actual household scanner panel data. In addition, I document a systematic relationship between the loyalty coefficient and the loyalty smoothing parameter. Insight for this systematic relationship and the multiple maxima is obtained by recognizing a trade-off between capturing household heterogeneity and state dependence. Finally, in the Guadagni–Little model extreme parameter values capture two different idealized forms of consumer behavior. However, reported studies rarely find these extreme parameter values. I show that procedures commonly used to initialize loyalty biases against these extreme parameter values. This bias offers some explanation for the observed empirical regularity in Guadagni–Little parameter estimates and suggests that researchers should be cautious concluding these parameters capture regularity in consumer behavior.

1. Introduction

Despite many advances in marketing models, the model proposed by Guadagni and Little (1983, henceforth GL) continues to be widely used among marketing science practitioners and academics. The longevity and widespread use are primarily due to the model's ability to fit scanner data extremely well with a small number of parameters. Because of its outstanding fit, the model is often used as a benchmark for improvements in marketing models. For example, Keane (1997) includes the GL loyalty variable in his model to examine state dependence in scanner data in ketchup. Bell and Lattin (2000) use the GL model in a study of loss aversion and reference price effects. Recently, Abramason et al. (2000) demonstrate that the GL model is quite robust against parameter bias from unobserved effects. The parsimonious nature of the model lends itself to practical applications and variants of the original GL model have been used at firms like Information Resources, Inc. for many years. Thus the model has been widely replicated in academic papers and applied in practice—two accomplishments most marketing models cannot claim.

In this paper I document two properties of the GL model. First, I show that the GL model can have multiple maxima. Second, I show that "standard" techniques used to initialize household loyalty tend to bias the GL smoothing parameter away from extreme values. This may be a contributing factor to the regularity in reported values of the GL smoothing parameter. These properties should be of interest to both academics and practitioners as I document them in simulated and actual scanner panel data.

The key feature of the GL model is the loyalty variable, which is specified as an exponential smoothing of past household purchases. As Guadagni and Little recognize in their original paper, the loyalty variable confounds two effects: state dependence (sometimes called purchase feedback) and household heterogeneity. Subsequently many authors have explored these two issues in depth (Heckman 1991; Vilcassim and Jain 1991; Fader and Lattin 1993; Gonul and Srinivasan 1993; Jaggia and Trivedi 1994; Kamakura et al. 1996; Papatla 1996; Roy et al. 1996; Bockholt and Dillon 1997; Gupta et al. 1997; Jedidi et al. 1997; Keane 1997; Allenby et al. 1998). Heckman (1991) argues that given sufficient flexibility in specification, models with heterogeneity and state dependence are not identified. GL recognize that their model specification adds parsimony at the expense of confounding state dependence and heterogeneity.

Importantly, presence of multiple maxima is in part explained by how the GL model treats state dependence and heterogeneity. The GL loyalty specification results in a systematic, non-linear relationship between the smoothing parameter and the loyalty coefficient. As the smoothing parameter increases, the loyalty coefficient also increases. Intuitively, one might describe two of the maxima that appear as either predominantly capturing heterogeneity or state dependence. One of the maxima has a smaller smoothing parameter and smaller loyalty coefficient, which implies more weight is placed on past purchases (state dependence). A second maximum has a larger smoothing parameter and larger loyalty coefficient, which places less weight on past purchases (heterogeneity).

This paper is organized as follows. In Section 2, I briefly summarize the GL model and describe the loyalty variable. In Section 3, I demonstrate that the likelihood function can have multiple maxima. I also show that there is a systematic relationship between the smoothing parameter and the loyalty coefficient. In Section 4, I explain why the GL model may not recover extreme parameter values and propose ways to make this more likely. I conclude with a brief discussion.

2. The GL Model

In this section, I briefly review the GL logit model and the loyalty variable. The GL model assumes the utility of a household h for brandsize j at purchase occasion t is $U_{h,j,t} = V_{h,j,t} + \varepsilon_{h,j,t}$. The term $V_{h,j,t}$ is the deterministic portion of utility and includes observable variables such as price, advertising, demographics, or past purchases, along with dummy variables for the alternative specific constants. Assuming an i.i.d. extreme value distribution for the error term leads to the familiar logit model. The probability of a household buying the jth brand at time t is given by

$$P_{h,j,t} = \frac{\exp(V_{h,j,t})}{\sum_{k=1}^{J} \exp(V_{h,k,t})}. \tag{1}$$

The likelihood function equals $\prod_h \prod_t \prod_j P_{h,j,t}^{y_{h,j,t}}$ where $y_{h,j,t}$ is an indicator variable equal to 1 if brandsize j was purchased at time t and 0 otherwise. The variable of interest in the GL model specification is loyalty. In the original paper, the authors consider both brand loyalty and size loyalty. For simplicity, I analyze a single brandsize loyalty variable. After dropping the household subscript, this becomes $\text{Loyalty}_{j,t} = \gamma\text{Loyalty}_{j,t-1} + (1 - \gamma)y_{j,t-1}$. The parameter γ is a smoothing parameter and $y_{j,t-1}$ is an indicator variable equal to 1 if brandsize j was purchased at time $(t-1)$ and 0 otherwise. Initial conditions for the loyalty variable are defined as $\text{Loyalty}_{j,1}$ where $\sum(\text{Loyalty}_{j,1}) = 1$. Re-writing the previous loyalty expression makes clear that the loyalty variable parsimoniously includes all past purchase indicators in the model.

$$\text{Loyalty}_{j,t} = \text{Loyalty}_{j,1} g^{t-1} + \sum_{k=1}^{t-1} g^{(t-1)-k}(1-g)y_{j,k}. \tag{2}$$

More generally, the GL model assumes that initial household preferences, lagged purchases, and current marketing influence the current purchase. The functional form of the GL loyalty variable is just one type of weighting of past purchases and initial conditions. To generalize, let the number of household purchases equal T_h. Deterministic utility for each household and brand can now be expressed as a function of the initial conditions, lagged purchases, and current marketing variables: $V_{h,j} = L_0\alpha + Y_{h,j}\theta + X_{h,j}\varphi$. In this expression, $V_{h,j}$ is a $T_h \times 1$ vector and L_0 and $Y_{h,j}$ are square matrices with dimension $T_h \times T_h$. The matrix L_0 contains initial conditions for loyalty and $Y_{h,j}$ lagged purchases. Both of these are illustrated in the Appendix. The term $X_{h,j}$ $(T_h \times K)$ captures the marketing mix variables and brand intercepts; α $(T_h \times 1), \theta$ $(T_h \times 1)$, and φ $(K \times 1)$ are parameters to be estimated[1]. The GL model defines the loyalty variable as an exponential smoothing of past purchases as in Equation (2). Mathematically, this places linear restrictions on the general model of the form $R\Delta = q$ where Δ is a vector of parameters (α and θ). After incorporating these linear restrictions, which are described in the appendix, the model reduces to $V_{h,j} = \text{Loyalty}_{h,j}\beta + X_{h,j}\varphi$. This notation will be useful in thinking about results from the next section where I show that the GL likelihood can have multiple maxima. One can interpret this result as stemming from the type of restrictions placed on the weight of past purchases and initial conditions. Whether other restrictions on these parameters lead to similar results is not explored in this paper. However, I note that consistent with Heckman (1991), some restrictions on the parameters θ and α may be necessary for identification.

3. Multiple Maxima

In this section, I provide examples to show that the GL model can have multiple maxima using both simulated and actual scanner data. I begin with a simulated example of one household and then consider an actual scanner panel dataset with multiple households and marketing mix variables. I demonstrate graphically that the likelihood, expressed as a function of β and γ, can have multiple maxima.

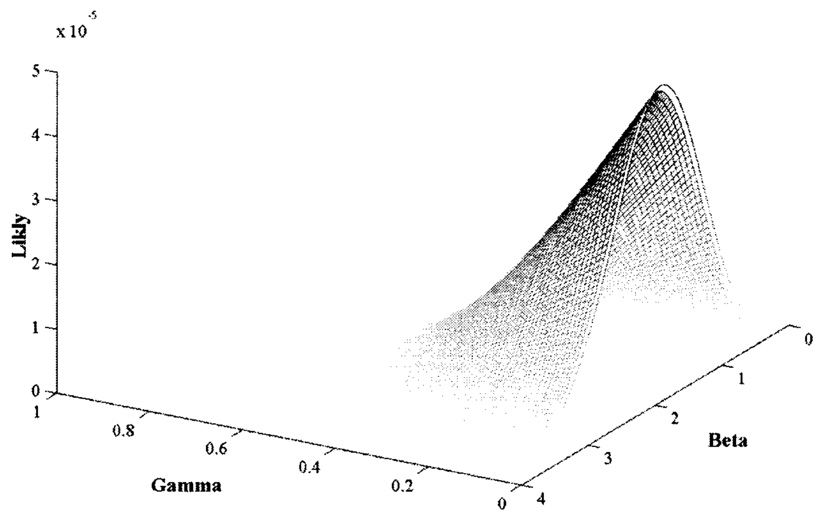
3.1. Example of Multiple Maxima

Consider a single household that makes 16 purchases as follows: 2 sequential purchases of brand C, followed by 4 sequential purchases of brand A, followed by 5 purchases of brand B, followed by 5 purchases of brand A. At the time of purchase, there are 3 available brands and loyalty is initialized at 0.33 (equal weights) for each brand. Figure 1(a) plots the GL likelihood for this household, which shows that the global maximum is at $\gamma = 0.00$ and $\beta = 1.93$ and the alternative specific constants are 1.10 and 0.58 for brands A and B, respectively. The log likelihood at this point is -9.92. The boundary solution is perhaps not surprising given the consistency in this household's purchases: the last purchase is a good indicator of the next brand purchase. Notice that no weight is placed on the initial conditions.

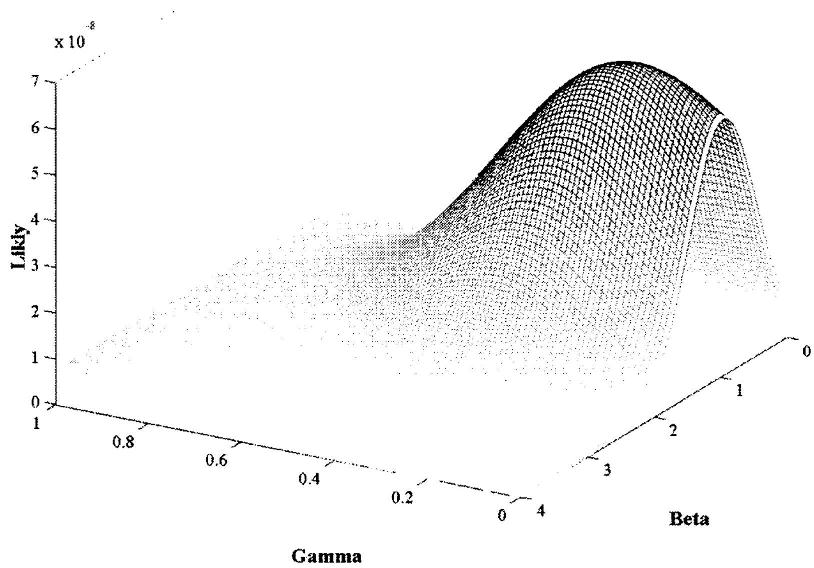
In Figure 1(b), I consider the same household's likelihood, but the 17th purchase is brand B and purchases 18–20 alternate between brands A and B. Now the global maximum is on the interior at $\gamma = 0.18$ and $\beta = 1.13$ and the alternative specific constants are 1.32 and 0.97 for brands A and B. The log likelihood at this point is -16.5. To accommodate the brand switching, the model places less weight on the last brand purchased as this is no long as good a predictor of future purchase.

To continue the example, I assume that on purchases 21–26 the household continues to alternate between brands B and A. The GL model places even less weight on past purchases, but now the likelihood takes on a unique shape. The global maximum is at $\gamma = 0.39$ and $\beta = 0.63$ and alternative specific constants equal 1.69 and 1.42 for brands A and B. The log likelihood at this point is -22.82. However, a second maximum emerges at $\gamma = 0.995$ and $\beta = 9.05$ with a log likelihood of -23.3. Figure 1(c) illustrates the shape of the likelihood function and depicts the global maximum. To better illustrate the second maximum, figure 1(d) narrows to $\gamma > 0.90$. While the likelihood surface is quite flat, looking at the region where γ is between 0.98 and 1 and β is 10 demonstrates existence of a local maximum. Finally, one can extend this example so that eventually the global maximum is at $\gamma \cong 1$ and a large value of β^2.

While the likelihood function varies dramatically in Figures 1(a)–(c), the relationship between β and γ is similar in all figures. To clearly illustrate this relationship, I calculate the optimal value of β for every value of γ, or $\beta^*(\gamma)$. Each value of $\beta^*(\gamma)$ is unique, as it is well known that the likelihood is globally concave conditional on γ and therefore has a unique maximum. Figure 2 is a plot of these values for the example in Figure 1(c), however a similar relationship holds for Figures 1(a) and 1(b).



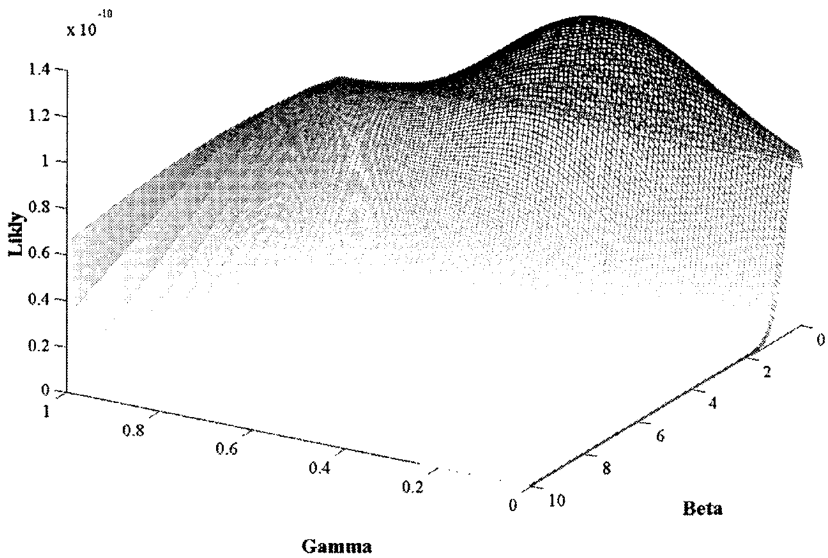
(a) GL Likelihood, Max at Boundary



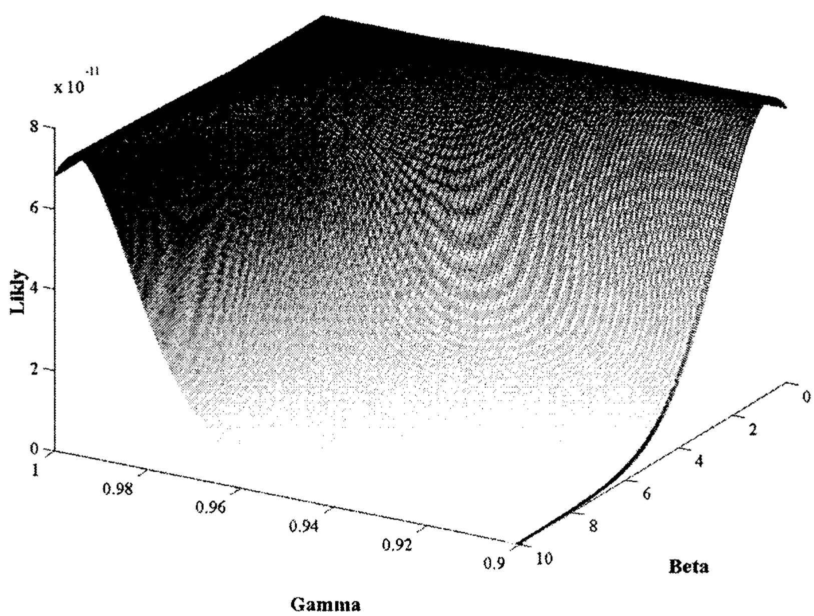
(b) GL Likelihood, Interior Global Maximum

Figure 1. Multiple Maxima Can Exist in GL Likelihood

The relationship between $\beta^*(\gamma)$ and γ shown in Figure 2 is quite robust and not specific to this example. I will utilize knowledge of this relationship to identify multiple maxima in a real-world example in the next section.



(c) GL Likelihood, Multiple Maxima



(d) GL Likelihood, Second Maxima

Figure 1. — continued

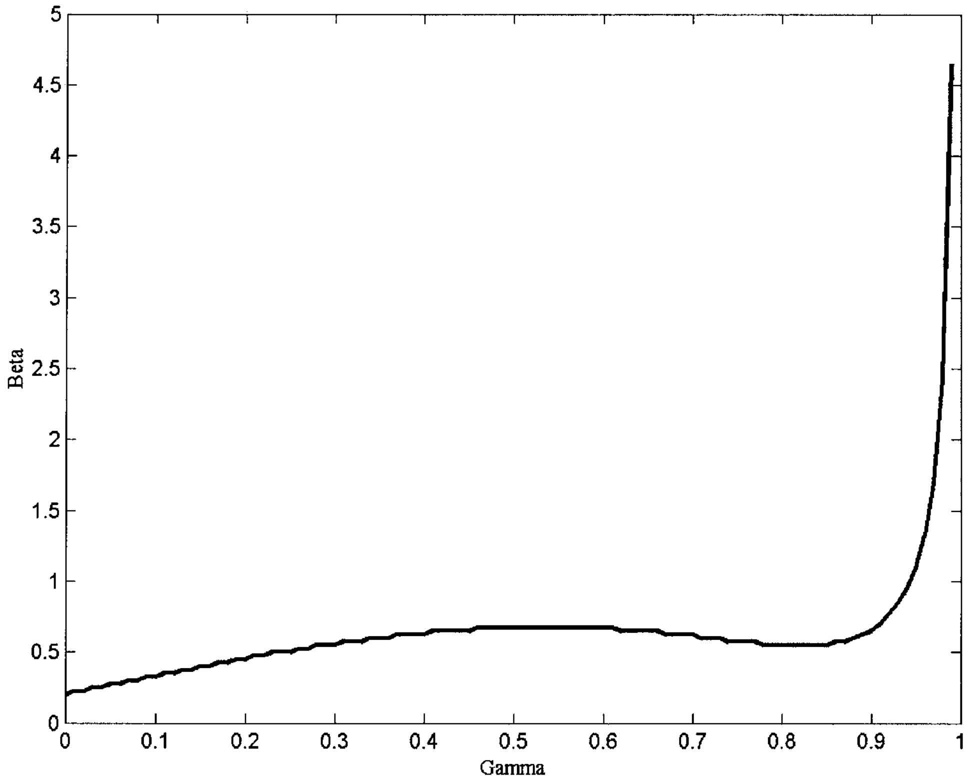


Figure 2. Optimal Beta Given Gamma.

3.2. Multiple Maxima in Scanner Panel Data

The previous examples considered the likelihood function for a single household using simulated data. An important question is whether the examples generalize to multiple households making actual purchases. In addition, the previous examples do not consider the impact of marketing mix variables. Will these examples survive when one includes these additional effects?

To investigate this issue, I estimate the GL model on scanner panel data from the detergent category. The particular dataset contains 10 brandsizes and 589 households that made 2,776 purchases over a two-year period. All customers purchased at least twice over the time period and households with extremely large numbers of purchases were dropped from the analysis. The 10 brandsizes were chosen based on unit share and represent over 50% of units sold. Loyalty was initialized using pre-period share of purchases for each household. The main result, the existence of multiple maxima, is not sensitive to how the data is cleaned. The model also includes alternative specific constants and the typical marketing mix variables: price per unit, display, and feature. In sum, this is a very typical scanner panel data set.

In Figure 3(a), I show the global maximum is at $\gamma = 0.65$, $\beta = 5.4$, which is a typical result for the GL model and was obtained using standard initial conditions and a gradient-based search method for estimation.[3] However, there is also another maximum value at $\gamma = 0.95$ and $\beta = 22.4$, which is plotted in Figure 3(b). At the global maximum, the log likelihood is -3071 while at the local maximum the log likelihood is -3238. I note that the likelihood has a similar shape to that illustrated in Figure 1(c) and there is a similar functional relationship between β and γ as shown in Figure 2. As one increases β and γ, the likelihood is not monotonically decreasing and a local maximum appears for large values of β and γ.

3.3. Multiple Maxima and Initial Conditions

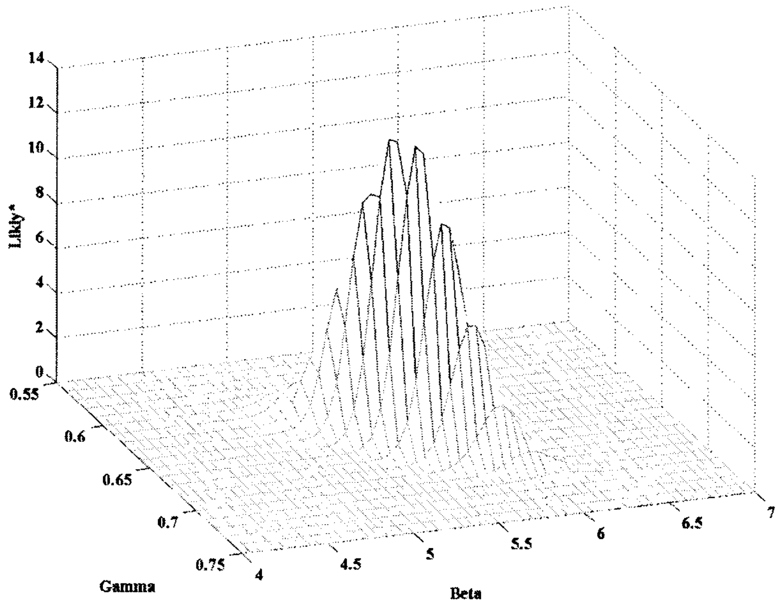
The previous examples document that the likelihood can have multiple maxima with respect to β and γ. In this section, I show that the presence of multiple maxima may extend to the relationship between γ and initial loyalty, $L_{j,1}$. One might argue that the GL model treats the initial conditions as exogenous parameters, and therefore the point is moot. However, I show that conditional on β and initial loyalty, the likelihood may have multiple maxima with respect to γ.

To illustrate this point, I set $\beta = 4$ and consider a market with 5 brands and a customer that purchases 2 units of brand A followed by 3 units of brand B. The initial loyalty of brands A and B is $L_{A,1}$ and $L_{B,1}$ and the initial loyalty of "other brands" equals 0.5. As shown in Figure 4(a), the likelihood has multiple maxima with respect to γ and $L_{A,1}$. When interpreting Figure 4(a), note that the range of "Initial Loyalty A" is affected by the loyalty of "other brands." Initial loyalty for other brands equals 0.5 and therefore initial loyalty of brand A ranges from 0 to 0.5. For example, a value of 0.20 for initial loyalty A implies that initial loyalty of brand B equals 0.30.

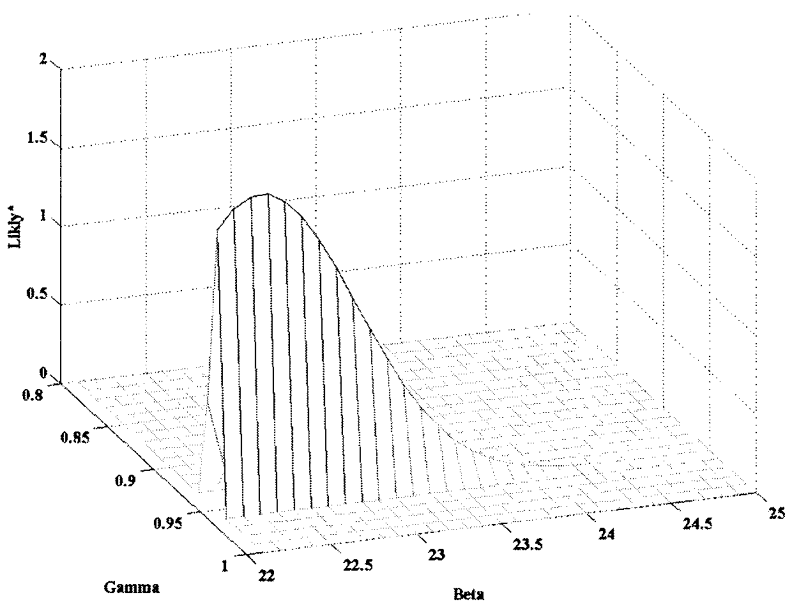
Conditional on β and $L_{A,1}$, the likelihood is also not well behaved with respect to γ. In Figure 4(b), I plot the likelihood for $\beta = 4$ and $L_{A,1} = 0.26$. Clearly this function has multiple maxima; there is a local maximum at the boundary ($\gamma = 0$) as well as an interior maximum.

3.4. Discussion of Multiple Maxima

The presence of multiple maxima is partially explained by how the GL model treats heterogeneity and state dependence. For example, in Figure 1(c) the global maximum is at low values of γ and β, which suggests a high degree of state dependence. In Figure 1(d), the global maximum is at high values of γ and β, which suggests past purchases have less impact and heterogeneity has a larger role. Figure 3 illustrates that a similar trade-off exists in actual scanner data. The confound between heterogeneity and state dependence offers some intuition for why multiple maxima are present in the GL likelihood.

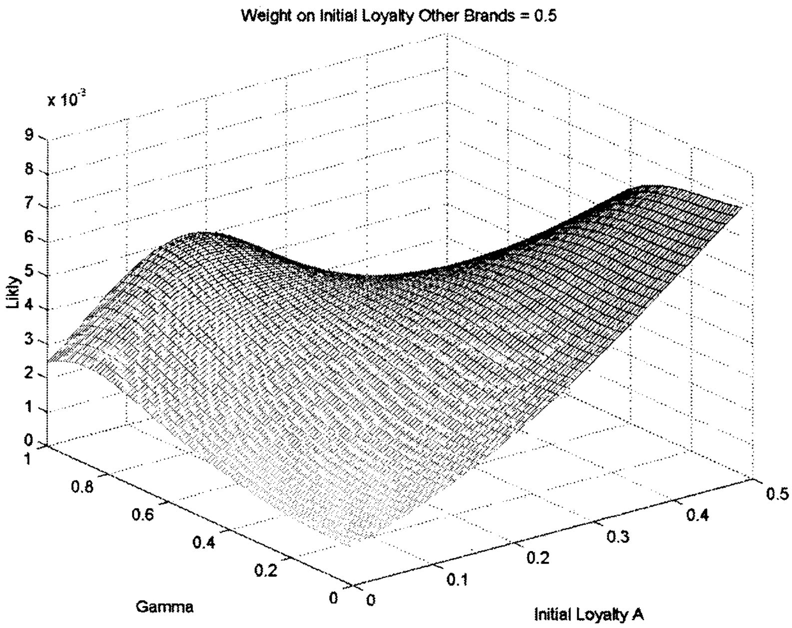


(a) Global Maximum at $\beta = 5.4$, $\gamma = 0.65$

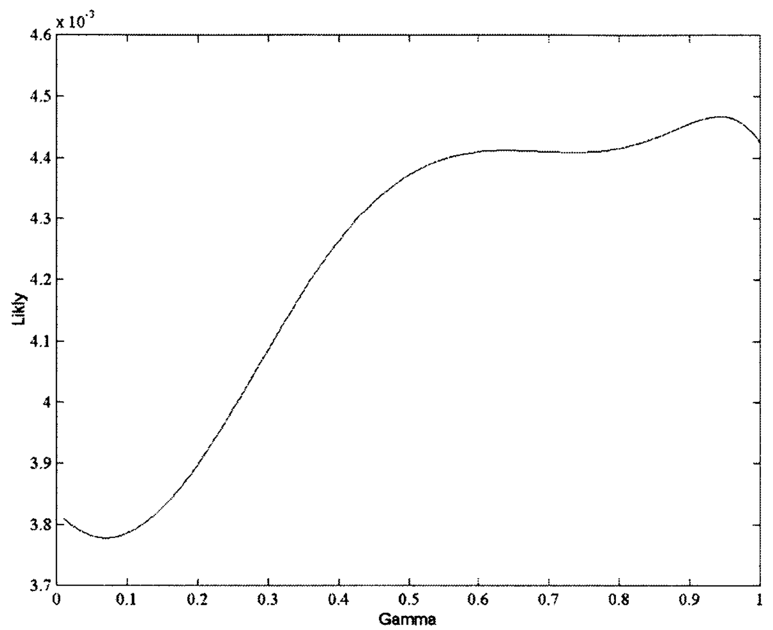


(b) Local Maximum at $\beta = 22.4$, $\gamma = 0.95$

Figure 3. Multiple Maxima in Household Detergent Data. (Note: Likelihood values scaled for computation purposes and should not be compared between figures 3(a) and 3(b).)



(a) Multiple Maxima



(b) Multiple Maxima given $L_{A,1} = 0.26$

Figure 4. Multiple Maxima in γ vs. Initial Loyalty.

4. GL Model and Consumer Behavior

In the previous examples, I show that maxima may occur at extreme parameter values ($\gamma \cong 1$ or $\gamma \cong 0$). In the GL model extreme parameter values capture two different idealized forms of consumer behavior. When $\gamma = 1$, the model posits no state dependence and static household preferences. In other words, a consumer's past purchases have no effect on future preferences. When $\gamma = 0$, the GL model captures a symmetric first order Markov model. To see this, let q equal the probability of re-purchasing a brand and $(1 - q)$ equal the probability of purchasing a different brand. In a first order Markov model, all information is contained in the last purchase and more distant purchases contain no information. Setting marketing equal to zero ($X = 0$), the probability of purchasing the same brand is $q = e^{\beta}/(J - 1 + e^{\beta})^4$.

In this section, I explain why standard techniques used to initialize household loyalty favors interior solutions (i.e. $\gamma \in (0, 1)$). I then suggest alternative techniques that model builders may want to consider. My approach is similar to Abramson et al. (2000), who simulated numerous datasets to explore properties of various logit models, including the GL model at three different values of γ. The authors explore bias in the loyalty coefficient (β) and other parameters, but treat γ as an exogenous parameter.

4.1. Problem: Finding Extreme Parameter Values

The GL model has been replicated in hundreds of published papers and the empirical regularity in the smoothing parameter is somewhat striking, (i.e. typically $\gamma \in [0.6, 0.8]$). I claim that at least one contributing factor is the "standard" method of initializing household loyalty. In the GL paper and subsequent replications, a sample of up to 12 months of data is set aside and used as a pre-period. A common method is to pick a starting value for γ, such as 0.7, and then create a loyalty vector for each household over the pre-period. Each household's final loyalty vector in the pre-period is then used as the initial value for the calibration period. The calibration data is used to estimate the parameters, including a new value of γ. To maintain consistency, one might iterate over γ until the pre-period and calibration values are equal.

One implication of this procedure is that every household's initial loyalty is non-zero for all brands. However, if the true value of γ is 0 then a household's loyalty should be 1 for the last brand purchased and 0 for all other brands. In Section 4.2, I show how this initialization may bias one against finding $\gamma = 0$ and propose a possible solution.

A second implication of the loyalty initialization technique is that using the same γ for the pre-period and calibration period results in a downward bias in γ. To see this, begin with the null hypothesis that there is no state dependence (H0: $\gamma = 1$). If one were to set $\gamma = 1$ for the pre-period, one would learn nothing about household preferences and be unable to control for heterogeneity. To control for heterogeneity using the pre-period data requires $\gamma < 1$. Thus, the "standard" initialization strategy is biased towards rejecting the null hypothesis. Using the same γ in the pre-period and calibration period ignores the motivation for having a pre-period, which is to learn about household preference and control for heterogeneity If one uses the same γ in both the pre-period and calibration period one is almost certain to find $\gamma < 1$.

4.2. *Solutions: Alternative Strategies for Initializing Loyalty*

The previous discussion suggests that one may need to consider alternative strategies for initializing household loyalty. To test for a first order Markov process, a simple solution is to use a pre-period of one purchase and to initialize each household with a loyalty vector equal to last brand purchased. To explore whether it is feasible to recover $\gamma = 0$ with this strategy, I simulate data with a first order Markov process and then estimate the GL model using this data. I simulated data for 200 households and varied the degree of state dependence ($q = 0.6, 0.7, 0.8$), the number of purchases (8 or 15), and the number of brands (6 or 15). I use two different procedures to initialize loyalty. First, I initialize each household's loyalty at the last brand purchased (i.e. true value). For the brand purchased in period 0, $L_{j,1} = 1$ and equals zero otherwise. This removes learning about household preference from the model but results in an initial loyalty that is degenerate (0/1). In the second procedure I again initialize loyalty with the correct state, but now use a pre-period as in the normal GL model. One might interpret this as introducing noise into the model. Results from these simulations show that when the initial conditions are degenerate and correct, the model performs quite well. The model consistently converged to estimates of γ and β that were both not statistically different from the true parameter values ($p < 0.05$). In contrast, when the initial conditions were not correct the model performed quite poorly. Estimates of γ were statistically different from zero in over half of the cases.

At the other extreme, differences in consumer behavior may be entirely due to heterogeneity. To enable the model to capture heterogeneity via the pre-period and still allow for $\gamma = 1$ in the calibration period, I consider a model with two smoothing parameters: γ_{pre} and γ_{cal}. The first parameter is used over the pre-period and the second parameter is used over the calibration period. This has the further advantage that the model now uses all the available data, which overcomes another criticism of the "standard" initialization technique. Both parameters can be estimated simultaneously, along with the marketing mix parameters, and the model uses all available information. A limit of this approach, and the GL model, is that the pre-period is arbitrary. The length of pre-period does effect the amount of information and hence the precision of γ_{pre} and γ_{cal}.

To demonstrate the feasibility of this technique, I simulated purchases for 200 households and varied the number of purchases (8 or 15), the number of brands available (6 or 15), the number of purchases in the pre-period (3 or 6), the GL smoothing parameter $(0.5, 0.7, 0.9, 1.0)$ and the loyalty coefficient $(1, 3, 4)$. Initial conditions for each household are set to $1/J$ and within a condition, each household had the same parameter values. The resulting purchase sequences for each household were not identical as they were simulated independently. For a given set of data, I estimated two models. The first model was the GL model and the second was the modified GL model that allows for two smoothing parameters, γ_{pre} and γ_{cal}.

The results show that when the true smoothing parameter is $\gamma = 1$, the average of the GL smoothing parameter is 0.84, which is significantly less than 1 ($p < 0.05$). The initialization clearly biases the GL smoothing parameter downward when the true value of γ is near 1. In contrast, when both γ_{pre} and γ_{cal} are estimated, the average estimate of γ_{cal} is 0.88. While γ_{cal} is closer to 1, the parameter is still biased. As the true value of γ decreases, the restriction $\gamma_{\text{pre}} = \gamma_{\text{cal}}$ is less severe. For example, if the true value is $\gamma = 0.70$ the average smoothing parameter in the GL model and the modified model are both 0.70. Thus, the GL initialization technique may not introduce bias when the true value of the smoothing parameter is less than 1.

While this solution is somewhat ad-hoc, it illustrates that the general solution to this problem is to specify a model with better controls for heterogeneity. This is well illustrated by Keane (1997) who shows that γ increases as controls for heterogeneity are added to the model. In other words, failing to control for heterogeneity results in a systematic downward bias in γ. Consistent with results in this paper, Keane (1997) finds an extremely large smoothing parameter (i.e. $\gamma > 0.9$) after controlling for various types of heterogeneity.

4.3. Discussion

The empirical regularity of the smoothing parameter in the GL model is well known. One possible explanation for this is that the GL model captures the behavior of an "average" household. Hence, the smoothing parameter tends to be moderate rather than extreme. Alternatively, the regularity in the smoothing parameter might be interpreted as capturing regularity in consumer behavior. However, the previous discussion suggests that one should be cautious applying this interpretation as the results show that estimates of the smoothing parameter are biased away from both $\gamma = 0$ and $\gamma = 1$.

For modelers, considering extreme values of γ is important for two very different reasons. First, these represent idealized consumer behaviors that researchers and practitioners may want to understand. Second, these extreme parameter values may also be local or global maxima. While it may be feasible to recover these extreme values in the GL model using "standard" initialization techniques for loyalty, one is very unlikely to do so.

5. Discussion and Conclusion

The analysis in this paper identifies several significant issues with the GL model.

- The GL likelihood function can contain multiple maxima.
- There is a systematic, non-linear relationship between β and γ.
- The GL model is theoretically flexible, but biased against finding extreme parameter values.

In the original paper, GL note that their model specification confounds heterogeneity and state dependence. The presence of multiple maxima and the systematic relationship between γ and β are both related to this confound. The results suggest that perhaps other models that are in the spirit of GL may have better theoretical properties. For example, some researchers use a static loyalty variable to capture heterogeneity and a last brand purchased indicator variable to account for state dependence. Note that this approach is simply a different parameter restriction on past purchases (see Appendix) and one may want to consider variants of this approach.

This analysis also sheds some light on the empirical regularity of the GL smoothing parameter. In most datasets, the GL smoothing parameter is between 0.6 and 0.8. Is this regularity reflecting some stability in customer behavior? Or is it a property of the model? Examples in this paper suggest that this regularity may in part be explained by the model specification, techniques used to initialize loyalty, and difficulty in capturing heterogeneity.

In sum, the paper illustrates important properties of the GL model of relevance to academics and practitioners. As the likelihood can contain multiple maxima, researchers and practitioners who use the model will want to ensure they have found the global maximum. Further, as the smoothing parameter is biased away from extreme values, one should be cautious offering a behavioral interpretation of the GL parameters.

Appendix: Implicit Parameter Restrictions in GL Model

The utility of customer h for brand j can be written as

$$V_{h,j} = L_0\alpha + Y_{h,j}\theta + X_{h,j}\phi \tag{A.1}$$

where L_0 is initial loyalty, $Y_{h,j}$ are past purchases, and $X_{h,j}$ is current marketing. Without any restrictions on the parameters, it is well known that the specification in A.1 is globally concave. However, the GL model places restrictions on the parameters in A.1 To illustrate this, consider a household with four purchases. The initial conditions matrix and weights are:

$$L_0 = \begin{pmatrix} L_{j,1} & 0 & 0 & 0 \\ 0 & L_{j,1} & 0 & 0 \\ 0 & 0 & L_{j,1} & 0 \\ 0 & 0 & 0 & L_{j,1} \end{pmatrix}, \qquad \alpha = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}.$$

Past purchases and weights are given by

$$Y_{h,j} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ Y_{j,1} & 0 & 0 & 0 \\ Y_{j,1} & Y_{j,1} & 0 & 0 \\ Y_{j,1} & Y_{j,1} & Y_{j,1} & 0 \end{pmatrix}, \qquad \theta = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}.$$

The restrictions on α and θ are given by $R\Delta = q$ where:

$$R = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ g & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & g & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & g & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g & -1 \end{pmatrix},$$

$$\Delta = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix}, \qquad q = \begin{pmatrix} -b \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Given these restrictions, the loyalty of household h for brand j equals $\beta \text{Loyalty}_{j,h} = L_0\alpha + Y_{h,j}\theta$.

To illustrate the restrictions consider the first two purchases. The restrictions imply that $(-\alpha_1) = (-\beta)$. On the first purchase, $L_0\alpha(t = 1) = L_{j,1}\alpha_1 = \beta L_{j,1}$. Because there are no lagged purchases at $t = 1$, $Y_{h,j}\theta(t = 1) = 0$. Thus, the loyalty term at $t = 1$ is simply $\beta L_{j,1}$. The restrictions also require $\alpha_1\gamma - \alpha_2 = 0$ or $\alpha_2 = \beta_\gamma$. In addition, $\theta_1 = \alpha_1 - \alpha_2 = \beta(1 - \gamma)$. At the second purchase, $L_0\alpha(t = 2) = L_{j,1}\alpha_2 = \beta_\gamma L_{j,1}$ and $Y_{h,j}\theta(t = 1) = Y_{j,1}\theta_1 = Y_{j,1}\beta(1 - \gamma)$. Thus, loyalty at $t = 2$ is $\beta[\gamma L_{j,1} + Y_{j,1}(1 - \gamma)]$.

Notes

1. For the brand intercepts (alternative specific constants), $X_{h,j}$ can be specified as a 0/1 dummy variable.
2. One can show that there exist values of β and γ such that at least one of the characteristic roots is positive and therefore the Hessian is not negative definite. This proves that the likelihood is not globally concave, but does not prove that there are multiple maxima.
3. In this case, loyalty was initialized as 0.75 and β was initialized at 3.0.
4. Solving this for β, one finds $\beta = \ln(J - 1) + \ln(q/(1 - q))$. When q equals 0 or 1, β is not identified since there is no variance in the data.

References

Abramason, Charles, Andrews, Rick L., Currim, Imran S., and Jones, Morgan. (2000). "Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice," *Journal of Marketing Research*, 37(4), 410–426.

Allenby, Greg M., Arora, Neeraj, and Ginter, James L. (1998). "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35(3), 384–389.

Bell, David, and Lattin, James. (2000). "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity," *Marketing Science*, 19(2), 185–200.

Ben-Akiva, Moshe, and Lerman, Steven R. (1985). *Discrete Choice Analysis*. Cambridge, MA: The MIT Press.

Bockholt, Ulf, and Dillon, William R. (1997). "Some New Methods for an Old Problem: Modeling Preference Changes and Competitive Market Structures in Pretest Market Data," *Journal of Marketing Research*, 34(1), 130–142.

Fader, Peter S., and Lattin, James M. (1993). "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12(3), 304–317.

Gonul, Fusun, and Srinivasan, Kannan. (1993). "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science*, 12(3), 213–229.

Guadagni, Peter M., and Little, John D.C. (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 1(2), 203–238.

Gupta, Sachin, Chintagunta, Pradeep K., and Wittink, Dick R. (1997). "Household Heterogeneity and State Dependence in a Model Purchase Strings: Empirical Results and Managerial Implications." *International Journal of Research in Marketing*, 14(4), 341–357.

Heckman, James J. (1991). "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, 81(2), 75–79.

Jaggia, Sanjiv, and Trivedi, Pravin K. (1994). "Joint and Separate Score Tests for State Dependence and Unobserved Heterogeneity," *Journal of Econometrics*, 60(1), 273–291.

Jedidi, Kamel, Jagpal, Harsharanjeet S., and DeSarbo, Wayne S. (1997). "Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity," *Marketing Science*, 16(1), 39–59.

Kamakura, Wagner A., Kim, Byung-Do, and Lee, Jonathan. (1996). "Modeling Preference and Structural Heterogeneity in Consumer Choice," *Marketing Science*, 15(2), 152–172.

Keane, Michael P. (1997). "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business & Economic Statistics*, 15(3), 310–327.

Papatla, Purushottam. (1996), "A Multiplicative Fixed-Effects Model of Consumer Choice," *Marketing Science*, 15(3), 243–261.

Roy, Rishin, Chintagunta, Pradeep K. and Haldar, Sudeep. (1996). "A Framework for Investigating Habits, "the Hand of the Past," and Heterogeneity in Dynamic Brand Choice," *Marketing Science*, 15(3), 280–299.

Vilcassim, Naufel J., and Jain, Dipak C. (1991). "Modeling Purchase-Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity," *Journal of Marketing Research*, 28(1), 29–41.