# A Consistent Loyalty Measure for Generalized Logit Models

FRED M. FEINBERG AND GARY J. RUSSELL

May, 2004

# A Consistent Loyalty Measure for Generalized Logit Models

## Abstract

Since the pioneering work of Guadagni and Little (1983), the most common method of accounting for household-level preference heterogeneity in the multinomial logit model has been the loyalty variable, an exponentially-weighted average of a household's past purchases. Although this approach has considerable intuitive appeal, it has not been provided with a rigorous statistical foundation. The present paper argues that loyalty variables, construed purely as cross-sectional measures of preference heterogeneity, should be constructed with reference to a particular model of choice behavior and assumptions regarding how marketing activity evolves over time. Beginning with a generalized form of the logit model consistent with both Jeuland's (1979) inertial choice framework and the Lightning Bolt model of Roy, Chintagunta and Haldar (1996), a loyalty measure is derived which is a consistent estimate of the household's underlying brand preferences; the proposed procedure is compared to other methods, including that of Guadagni and Little, using both simulated and actual household-level scanner panel purchase data. Relative to the derived loyalty measure, estimated choice models based upon these alternative measures fit more poorly and, intriguingly, show an inability to account for data with a pronounced first-order choice dependence. We argue that the proposed loyalty measure's superior performance owes to its derivation from a modelling framework capable of distinguishing the effects of habit-persistence and brand loyalty.

# 1 Introduction

Among the most durable choice models in marketing is the multinomial logit (MNL) model. Introduced to the marketing literature by Guadagni and Little (1983), the MNL model has been applied in a wide variety of contexts to study the impacts of price, promotion, advertising and other managerial control variables in consumer panel data (e.g., Carpenter and Lehmann 1985, Lattin and Bucklin 1989, Anderson 2002). A key issue in implementing the model is proper control for preference heterogeneity across households. It is has been well-known for quite some time that incorrect assumptions regarding household-level heterogeneity can seriously bias estimates of the impact of marketing mix elements upon brand choice (Gonul and Srinivasan 1993). Since then, numerous papers have successfully applied the Guadagni-Little loyalty measure – for example, Bell and Lattin (2000) to study loss aversion and reference pricing, Abe et al. (2003) in competitive pricing, Seetharaman (2003) for state dependence – and it has been found not only to perform well, but to be relatively robust in terms of parameter bias (Abram et al. 2000). Because the measure is so easily applied, it will certainly continue to find wide use, and so a deeper understanding of its statistical properties is called for.

## 1.1 Models of Heterogeneity

Work on heterogeneity correction in logit models has developed along two parallel paths. The older, classical path takes the point of view that individual effects can be estimated at the household level, and is in its way reminiscent of approaches common in the literature on stochastic choice. Jones and Landwehr (1988), for example, regard individual preferences as "nuisance parameters" which can be removed from the logit model likelihood function by conditioning on certain household purchase history statistics. Guadagni and Little (1983) proposed that an exponentially-smoothed average of past purchases be included in the model specification; this so-called loyalty variable is designed to correct for both cross-sectional (preference) heterogeneity and for nonstationarity in preferences over time. More recent work has attempted to rid the basic Guadagni-Little (GL) loyalty variable of contamination by marketing mix elements (Allenby and Rossi 1991) and to allow for renewals in brand preference values over time (Fader and Lattin 1993). The alternative research path views consumer heterogeneity as an unobserved cross-sectional effect which must be estimated along with

the marketing mix coefficients of interest. This effect can be represented using random coefficient models (e.g., Chintagunta, Jain and Vilcassim 1991, Gonul and Srinivasan 1993) or fixed coefficient mixture models (e.g., Kamakura and Russell 1989). More complex heterogeneity representations can be constructed using Hierarchical Bayes logic and analyzing an augmented logit model likelihood using Markov Chain Monte Carlo techniques (e.g., Allenby and Lenk 1994). A common characteristic of these approaches is the assumption that data are too sparse for individual estimation of brand preferences so that, in effect, such estimates are formed by a stochastic pooling of data across households. We acknowledge the superiority of current Bayesian methods (e.g., Rossi and Allenby 2003) in accounting for complex coefficient and error specifications. However, a major attraction of loyalty variables is to sidestep the complexity and computational burden such methods require. A deeper discussion of these issues, methods of loyalty variable construction and associated optimization, is provided by Anderson (2002) and by Seetharaman (2003).

## 1.2 Loyalty Variables

In the present paper, we build upon the classical methods of heterogeneity correction by developing an approach for constructing loyalty variables in logit choice models. This work is motivated by two considerations. The loyalty variable approach is extensively used by both the market research community (e.g., IRI and A.C. Nielsen) and the marketing academic community (e.g., Tellis 1988, Lattin and Bucklin 1989, Allenby and Rossi 1991). Although the fixed and random coefficient approaches possess desirable properties, loyalty variables exert a strong intuitive appeal, are straightforward to create, and can be used with standard logit software. Simply put, loyalty variables are used because they are an easily-applied method for substantially improving choice model forecast accuracy. Unfortunately, despite their active use, loyalty variables lack a rigorous theoretical basis, and thus little is known regarding the appropriate circumstances under which they might be applied or the biases, if any, they may introduce. With the notable exception of that of Jones and Landwehr (1988), classical heterogeneity methodologies have not been derived from a statistical analysis of an assumed underlying choice mechanism. Because loyalty variables are based upon observed choices, many researchers have argued that loyalty variables are contaminated by marketing mix activity and so cannot be regarded as proper estimates of a household's true brand preferences; for example, during a period of frequent promotional activity, loyalty variables will tend to favor the promoted

4

brand. Neither Guadagni and Little (1983) nor Fader and Lattin (1993) fully address this issue.

## 1.3   Research Agenda

The present research is carried out under the supposition that loyalty variables will continue to exert influence, for the reasons cited above, in a variety of applied choice modelling contexts. Accordingly, our goal is the rigorous derivation of a loyalty measure suitable for application in logit models for the analysis of scanner panel and web-based choice data. We begin by conceptually defining a loyalty variable as a consistent estimate of parameters measuring a household's intrinsic brand preferences. This definition is then applied in the context of the Inertial Logit model, a generalized version of the standard logit formulation incorporating Jeuland's (1979) model of brand choice inertia, which is generalized in a dynamic framework by the Lightning Bolt model of Roy, Chintagunta and Haldar (1996). When marketing mix activity can be taken to induce mean stationary choice probabilities over time, an appropriate (i.e., consistent) loyalty variable for the Inertial Logit model can be derived. Finally, using both simulated and actual consumer choice histories, we argue that the proposed loyalty measure provides a valid method for capturing preference heterogeneity in the choice process. This approach yields two considerable insights. First, an explicit decomposition is obtained whereby the effects of marketing mix activity on loyalty variables are clarified; in general, such activity will introduce bias into the loyalty variable, but the resulting bias can be mitigated under appropriate circumstances. Second, we show that loyalty measures of the type derived only 'work' under certain assumptions about the evolution of both consumer choice behavior and marketing mix activity. Thus, the present research suggests conditions under which loyalty variables might be less appropriate than alternative methods of heterogeneity correction.

At the outset, it must be stressed that the term 'loyalty' has been applied in many contexts in the Marketing field at large, and no generally agreed-upon definition can be easily summoned up. While household-level purchase behavior and temporal marketing-mix trends are part and parcel of any discussion of the locus of applicability of such measures, no attempt is made to offer a definition of the general phenomenon of loyalty as the term is applied in common parlance. The present study employs this term in a very specific sense, namely, to refer to a consistent measure of household-level preference heterogeneity, specifically as applied in the context of logistic regression as a tool for the analysis of scanner panel data.  For a full discussion of loyalty variables and their connections to

habit persistence and state dependence, see Seetharaman (2003).

## 2 Defining the Logit Loyalty Variable

One of the essential steps in the development of loyalty variables is understanding the role they assume in the analysis of choice data. In this work, we take an explicitly statistical perspective and regard loyalty variables as estimates of certain parameters of the choice model. Viewed in this fashion, loyalty variables cannot be constructed unless the researcher specifies both the choice model under consideration and the parameters which the loyalty variables are intended to estimate. Below, we detail these selections for a logit-based choice process that can account for the effects of habit-persistence in the sense of Heckman (1981).

### 2.1 Inertial Logit Model

The analysis to be presented here is based upon a generalization of the multinomial logit model to allow for choice-based feedback (habit-persistence, or inertia), as first put forth by Roy, Chintagunta and Haldar (1996). Let $k$ denote a particular household and define the household purchase indicator variables:

$$Y_{ikt} = \begin{cases} 1, & \text{if household } k \text{ chooses brand } i \text{ at time } t \\ 0, & \text{otherwise} \end{cases}$$

with respect to brands $i$ and time points $t$. Then, purchase probabilities for household $k$ are given by:

$$P_{ikt} = JY_{ik(t-1)} + (1-J)\theta_{ikt} \tag{1}$$

where $J \in [0,1]$ is a parameter measuring the quantity of inertia in the choice process. The Inertial Logit model takes the observed choice probability $P_{ikt}$ at time $t$ to be a weighted average of the directly prior choice and the current-period latent choice probability:

$$\theta_{ikt} = \frac{A_{ikt}}{\sum_j A_{jkt}} \tag{2}$$

in multinomial logit form. The brand attractions are given by the standard linear specification:

$$A_{jkt} = \exp\left[\alpha_{jk} + \sum_r \beta_r X_{rjt}\right] \tag{3}$$

where $X_{rjt}$ denote marketing mix variables (e.g., price and promotion). Thus, both prior choices and current marketing mix activity affect the household's current brand choice, the first in a Markovian fashion.

This model generalizes the logit model to include a specific type of choice inertia; by invoking parametric restriction $J = 0$, the model reduces to the standard logit model. Consistent with this standard model, the Inertial Logit model assumes that heterogeneity across consumers is restricted to the brand-specific intercepts $jk$. However, in contrast to Guadagni and Little (1983) and Fader and Lattin (1993), this model does not assume nonstationarity, that is, that the brand-specific intercepts drift over time. In contrast, the Inertial Logit model posits a specific mechanism, moderated by the value of $J$, which relates the current choice probability to prior choices. For values of the inertial parameter $J$ near one, the current choice probability is largely determined by the prior brand choice. The inertial process in this model can be viewed as a type of preference nonstationarity which depends upon the prior choice history.

The Inertial Logit model can also be viewed as an extension of Jeuland's (1979) first-order stochastic brand choice model to include marketing mix variables. This linkage to Jeuland's work leads to an important observation. Suppose that the marketing mix variables $X_{rjt}$ are unknown to the researcher, but that they vary over time in such a way that the latent choice probabilities $\theta_{ikt}$ follow a stationary stochastic process over time with a fixed mean; that is, treat the latent choice probabilities as random variables and suppose that the expectation, $E[\theta_{ikt}] = \theta_{ik}$, is independent of time. Then, as shown in the Appendix, the unconditional expectation of the household purchase indicator variables,

$$E[Y_{ikt}] = E[\theta_{ikt}] = \theta_{ik} \tag{4}$$

is independent of $J$, the quantity of choice inertia in the process. This remarkable property is related to that fact that the equilibrium point of Jeuland's stochastic process is also independent of the value of the choice inertia parameter. As will be seen subsequently, equation (4) facilitates both the construction of loyalty variables for this model and its estimation on household-level purchase history data.

## 2.2 Logit Loyalty Variable

As noted earlier, the concept of a loyalty variable has not been supplied with a definition which has achieved consistent use in the Marketing literature. Here, we follow the implicit definition of Fader and Lattin (1993), which stipulates that loyalty variables should recover a reasonable estimate of the logit model's brand-specific intercepts. Formally, then, we define loyalty variables in the following manner. Let

$$\pi_{ik} = F(Y_{ik1}, Y_{ik2}, ..., Y_{ikT}) \tag{5}$$

be some function of the household's entire choice history. Then, $\pi_{ik}$ is a *loyalty variable* for the Inertial Logit model if

$$plim \ \pi_{ik} = \alpha_{ik} \tag{6}$$

where *plim* denotes the probability limit as the length of the purchase history increases without bound. Put differently, a loyalty variable is taken to be a consistent estimate of a household's brand-specific intercept.

# 3  Loyalty Variable Specification

Although loyalty variables can be defined in principle, the definition (6) does not provide guidance as to their particular formal specification. Rather, the most prudent route available to the researcher is to postulate a particular form for the loyalty variable and to demonstrate that it possesses the desired consistency property. A key implication of this analysis is that the justification for loyalty variables, in the sense of preference heterogeneity correction, depends intrinsically upon both the structure of the choice model and upon the manner in which marketing mix activity evolves over time; this is the route taken below with regard to the probability specification (1).

## 3.1  Long-Run Choice Mean

We begin by considering the properties of the long-run choice mean

$$L_{ik} = \frac{1}{T_k} \sum_{t=1}^{T_k} Y_{ikt} \tag{7}$$

where $T_k$ is the length of the choice history of household $k$. $L_{ik}$ is simply the proportion of choices allocated to brand $i$ by the household, typically called the share of requirements for brand $i$ by the

marketing research community.

The $L_{ik}$ statistic is interesting for two reasons. First, it is perhaps the simplest and most intuitive measure of household preference which can be constructed, akin to the assumption of an unadorned multinomial choice process. Second, it underlies much of the previous work using loyalty variables, being a special case of the Guadagni and Little loyalty variable, as well as having been used as a heterogeneity measure in a variety of previous studies (e.g., Lattin and Bucklin 1989, Tellis 1988). Third, although it employs equal weighting of the prior purchase indicator variables, formal manipulations of the expression (7) can be directly extended to other weighting schemes, such as the more typical exponential smoothing. Hence, an understanding of the properties of the long-run choice mean sheds considerable light upon loyalty-type heterogeneity corrections appearing in previous research.

Notice that $L_{ik}$ obeys the general form of equation (5), being a function only of the observed purchase indicator variables for brand $i$. From a statistical perspective, this implies that the properties of the long-run choice mean need be derived without knowledge of the marketing mix activity which influenced the actual choices. Accordingly, $L_{ik}$ should be viewed as an estimate of the time average of $Y_{ikt}$ with respect to the unconditional temporal distribution of $Y_{ikt}$.

## 3.2 Choice Model Analysis

To examine the properties of $L_{ik}$, we first compute the value of $E[\theta_{ikt}]$, the period-by-period expected latent choice probabilities in the Inertial Logit model. Recall that these latent probabilities are defined as

$$\theta_{ikt} = \frac{A_{ikt}}{\sum_j A_{jkt}} \quad ; \qquad A_{jkt} = \exp\left[\alpha_{jk} + \sum_r \beta_r X_{rjt}\right] \tag{8}$$

where $X_{rjt}$ represent the marketing mix elements. For reasons which will become clear subsequently, the value of $E[\theta_{ikt}]$ leads directly to an expression for the household's long-run choice mean.

The evaluation of $E[\theta_{ikt}]$ calls upon two separate assumptions regarding the manner in which marketing mix variables evolve over time. The first of these temporal assumptions states that at any time point $t$, $\theta_{ikt}$ is independent of the total attraction $\sum_j A_{jkt}$ for brands in the category. While this is not an explicit assumption of attraction models as a class, it is nonetheless a reasonable one. Practically speaking, it presumes that, should the category in question become more or less attractive to the household *as a category*, relative preferences for the brands *within* that category are

9

not affected. Equivalently, this assumption implies that knowledge of the sum of brand attractions $\sum_j A_{jkt}$ for a particular consumer provides no information about the relative sizes of the individual $A_{jkt}$, up to multiplicative scaling. Note that this assumption does not imply that $\sum_j A_{jkt}$ is independent of $A_{jkt}$, nor does the reverse implication hold.[1]

This assumption has an important application in the problem at hand; because, $\theta_{ikt}$ and $\sum_j A_{jkt}$ are uncorrelated, we can write:

$$E\left[A_{ikt}\right] = E\left[\theta_{ikt}\sum_j A_{jkt}\right] = E\left[\theta_{ikt}\right]E\left[\sum_j A_{jkt}\right] = E\left[\theta_{ikt}\right]\sum_j E\left[A_{jkt}\right] \tag{9}$$

Rearranging this expression yields the useful result:

$$E\left[\theta_{ikt}\right] = \frac{E\left[A_{ikt}\right]}{\sum_j E\left[A_{jkt}\right]} \tag{10}$$

The second temporal assumption allows further analysis of the expected attraction terms in (10). We assume that the marketing effort for each brand can be decomposed into a seasonal component which is common across brands, and into a random component which is specific to the brand in question (as in Kim, Menzefricke and Feinberg 2003). Formally, this requires that the marketing mix impact variables for brand $i$ can be expressed as

$$Z_{it} = \sum_r \beta_r X_{rit} = \phi_t + \mu_{it} \tag{11}$$

where $\phi_t$ is independent of $\mu_{it}$. While $\phi_t$ is taken to be an arbitrary function of time, $\mu_{it}$ follows a stationary time series process, with a mean that is (potentially) non-zero and brand-specific.

The decomposition (11) allows the evaluation of the expected brand attraction at time $t$ as

$$E\left[A_{ikt}\right] = M_i \cdot E\left[\phi_t\right] \cdot \exp\left[\alpha_{ik}\right] \tag{12}$$

where $M_i$ is a constant specific to brand $i$, due to the stationarity of $\mu_{it}$. Consequently, at each time point $t$, the expected (latent) choice probability is

$$E\left[\theta_{ikt}\right] = \frac{E\left[A_{ikt}\right]}{\sum_j E\left[A_{jkt}\right]} = \frac{M_i \exp\left[\alpha_{ik}\right]}{\sum_j M_j \exp\left[\alpha_{jk}\right]} \tag{13}$$

---

[1]This is an important distinction which, in our experience, is often misunderstood. It is plainly incorrect to assume that, if the attraction of one brand receives a shock, that the total attraction of its category is unchanged. What we make use of here is different: that, if one observes a change in total category attraction, there is no reason to attribute it to any one brand disproportionately to its predicted choice share. It must be emphasized that the validity of equation (10) rests upon the assumption of independence between brand choice probabilities and total category attraction, but says nothing whatever about brand *attractions*.

which is independent of time. Notice that this expression depends upon both the brand marketing mix activity (through $M_i$) and the household-specific brand intercepts ($\alpha_{ik}$), and precisely quantifies the degree of anticipated 'bias' in the expected choice probabilities based solely on brand-specific terms.

## 3.3 Expectation of $L_{ik}$

These results allow calculation of a household's expected long-run choice share. Recall that, as demonstrated in the Appendix, when $E[\theta_{ikt}]$ is time-invariant, $E[Y_{ikt}] = E[\theta_{ikt}]$ in the Inertial Logit model. This result, coupled with (13), shows that

$$E[L_{ik}] = T_k^{-1} \sum_{t=1}^{T_k} E[Y_{ikt}] = T_k^{-1} \sum_{t=1}^{T_k} E[\theta_{ikt}] = \frac{M_i \exp[\alpha_{ik}]}{\sum_j M_j \exp[\alpha_{jk}]} \tag{14}$$

so that these expectations are independent of the degree of habit-persistence (choice inertia) in the process. Hence, given these two temporal assumptions, one obtains an explicit expression for the expected proportion of purchases allocated to each brand over the household's choice history.

Although the derivation of this relationship is somewhat complex, the expression for the expected long-run choice mean is quite amenable to interpretation. According to equation (14), vigorous marketing mix activity (as measured by large values of the brand constants $M_i$) upwardly biases the observed choice mean away from the true brand intercepts $\alpha_{ik}$. Thus, as many researchers have suggested, $L_{ik}$ is a biased measure of the intrinsic brand preference of the household; equation (14) simultaneously quantifies this bias and suggests how it might be corrected for.

This definition is motivated by the notion that loyalty variables can be used to estimate the parameters of the Inertial Logit model in a two-step process. First, one obtains a consistent estimate of the brand intercepts; second, by conditioning upon these estimates, the remaining model parameters are estimated. Previous work in Marketing employing loyalty variables in a classical setting has followed this prescription of estimating marketing mix variable coefficients conditional upon the computed values of loyalty variables (e.g., Lattin and Bucklin 1989). Among the advantages of this approach is a simplified estimation procedure for the analyst. As will be seen subsequently, this definition of loyalty is overly strict; a biased loyalty variable can nevertheless be of use in calibrating a choice model. However, this definition, despite its narrow focus, makes unambiguous the goals for loyalty variable construction; within the context of a specific choice model, the choice history is

called upon to provide a household-level estimate of brand intercepts. If such a estimator can be found, the resulting estimates are termed loyalty variables and used to define a two-step estimation approach.

## 3.4 Loyalty Variable Specification

Formally, it has been shown that $L_{ik}$ is not a loyalty variable by our definition, in that it is not a consistent estimate of the brand-specific constants $\alpha_{ik}$. Nevertheless, the preceding construction can be modified so as to create an asymptotically-correct loyalty measure based on $L_{ik}$. Notice that, although $L_{ik}$ is biased due to marketing mix activity, the biases depend on the brand only, *not* on the particular household. That is, the long-run choice means of all households are biased, but all means are biased in the same way, as quantified by (14).

This observation suggests the following procedure. Define $\text{LOY}_{ik} = \log(L_{ik})$ and note that the probability limit of this measure is given by

$$plim \ \text{LOY}_{ik} = \gamma_k + \log(M_i) + \alpha_{ik} \tag{15}$$

where $\gamma_k = \log\left[\sum_j M_j \exp\left(\alpha_{jk}\right)\right]$ is a constant independent of brand. Next, consider the Inertial Logit model with brand attractions defined as

$$\text{ATT}_{ikt} = \exp\left[\tau_i + \tau_0 \text{LOY}_{ik} + \sum_r \beta_r X_{rit}\right] \tag{16}$$

as opposed to the typically-used specification of equation (3). Notice that, if $\tau_i = -\log(M_i)$ and $\tau_0 = 1$, then $plim \ \text{ATT}_{ikt} = exp\left(\gamma_k\right) A_{jkt}$. Consequently, $plim \left[\text{ATT}_{ikt}/\sum_j \text{ATT}_{jkt}\right] = A_{ikt}/\sum_j A_{jkt} = \theta_{ikt}$, which implies that the model specification is asymptotically correct.

We refer to the methodology implied by equation (16) as a loyalty variable procedure. Using equation (16) instead of equation (3) to specify the Inertial Logit model, we compute $\text{LOY}_{ik}$ and then employ a maximum likelihood algorithm to estimate the remaining parameters (i.e., the $\tau_i$, $\tau_0$ and the $\beta_r$). Details of the estimation procedure are addressed below. Although $\log(L_{ik})$ does not satisfy the consistency-based definition for a loyalty variable, the bias in $\log(L_{ik})$ arising from marketing mix activity can be removed by estimating the common brand intercepts $\tau_i = -\log(M_i)$. Hence, the approach is functionally equivalent to a two-step estimation procedure in which true loyalty variables are available in the first stage.

## 3.5 Summary

The developments of this section provide a compelling illustration of the challenges facing the researcher working with the loyalty variable approach. Even though special properties of the Inertial Logit model were exploited, it was necessary to invoke several assumptions about the time evolution of marketing mix activity in order to justify the use of a loyalty variable procedure. We next turn to two empirical applications, one simulation-based, the other involving scanner panel data, to examine the performance-based properties of the proposed loyalty approach.

# 4  Simulation Study

In order to better understand aspects of the performance of the proposed loyalty measure, we examine it in a simulation-based setting. An appropriately-chosen set of simulation scenarios allows the assessment of statistical properties that may not be readily apparent within any group of data sets. We consider a two-brand market, with each brand equally-preferred at the category level. In the simulated data, all marketing mix activity is condensed into a single variable, which is termed, for simplicity, Price. These selections simplify the interpretation of the simulated data, but are rich enough to demonstrate the principal statistical properties of the proposed loyalty measure.

## 4.1  Structure of the Simulated Data

The simulated choice data are designed to capture three key elements typical of actual choice histories: heterogeneity in consumer preferences, feedback from previous brand purchases (habit-persistence), and autocorrelation in marketing mix activity (structural state-dependence). The simulated data are constructed according to the following procedure.

*Choice Process.* For simplicity, the market consists of two brands, and choice histories are simulated for 20 households. For a given value of $J$, the probability $P_{1kt}$ that consumer $k$ buys Brand 1 at time $t$ is given by an analog of the Lightning-Bolt (LB) model:

$$
\begin{aligned}
P_{ikt} &= JY_{1k(t-1)} + (1-J)\theta_{1kt} &\qquad (17)\\
\theta_{1kt} &= \left[1 + \exp\left(-\delta_k + \beta \cdot Price_t\right)\right]^{-1}
\end{aligned}
$$

*Autocorrelated Marketing Activity.* The Price time series (interpreted here as the difference in prices between brand 1 and brand 2) is a first-order autoregressive process with parameter $\rho$, again

13

consistent with the LB formulation:

$$Price_t = \rho \cdot Price_{(t-1)} + (1 - \rho) \cdot \epsilon_t \tag{18}$$

$$\epsilon_t \sim N[0, 1], \qquad Price_{(t=0)} = 0$$

*Preference Heterogeneity.* Across households, the baseline preferences for Brand 1 (measured as $[1 + \exp(-k)]^{-1}$) are evenly spaced over $[.2, .8]$. Within each household, baseline preferences are constant over time, that is, nonstationarity is not at issue.

*Simulation Parameters.* A factorial 2 x 2 design is used to vary price series autocorrelation ($\rho = \{.1, .5\}$) and the choice inertia parameter ($J = \{.1, .4\}$). In all cases, the price coefficient was set to $\beta = -2$. These simulation parameters have been chosen to emulate moderate or typical values in the literature; we ignore coefficient heterogeneity here to simplify interpretation of results. It is important to remember that Price in this two-brand setting represents the price difference between the brands; hence, autocorrelation in the simulated data refers to the tendency for such a difference to persist from one period to the next.

## 4.2    Model Specification and Benchmarks

There are two explicit goals in fitting the Inertial Logit model to data generated in accordance with its underlying structure. The first such goal involves understanding the degree to which the known parameters of the model can be recovered through standard econometric techniques, therefore being able to anticipate bias and quality of convergence. The second goal involves the structure of the Inertial Logit model itself. We wish to determine whether the purchase feedback structure in the Inertial Logit model (i.e., habit-persistence) can be adequately represented by more conventional logit model structures. While the first of these goals can be accomplished through simply estimating the Inertial Logit model on a variety of simulated data sets, the second requires the specification of alternative benchmark models. In this regard, two immediately suggest themselves: a non-inertial version of the Inertial Logit model and the model of Guadagni and Little (1983) (GL). If the Inertial Logit model is fit subject to the constraint that $J = 0$ (that is, presuming there is no inertia), we are left with a standard logit formulation making use of the simple-average household-loyalty measure. Because this model is nested within the Inertial Logit model, it provides a clear measure of the additional utility afforded by incorporating inertia. The GL model is an appropriate benchmark for

14

a number of reasons. The loyalty variable in GL makes use of a calibration period, and as such can be compared to the present model based on the same input data. The GL Loyalty measure for a given brand $i$ and household $k$ takes the form of an exponentially-smoothed variable

$$LOY_{ikt} = \lambda \cdot LOY_{ik(t-1)} + (1 - \lambda) \cdot Y_{ikt} \tag{19}$$

where $0 < \lambda < 1$. This measure is designed to account both for household heterogeneity and preference nonstationarity. Note that the GL model employs the same number of parameters as the Inertial Logit, requiring that a smoothing constant ($\lambda$) be estimated, rather than a carryover parameter ($J$).

These two benchmark models can be estimated through well-known techniques; the Simple Logit model with loyalty requires only standard logistic regression software, while the GL model can be estimated using an iterative modification of the standard logit estimation procedure developed by Fader, Lattin and Little (1994) for fitting non-linear parameters (such as $\lambda$) in the utility specification. The Inertial Logit model, on the other hand, cannot be fit into this framework because of the way in which choice inertia enters the model. Pursuant to this goal, we have developed a method which extends the Fader-Lattin-Little technique to account for state-dependent specifications, that is, ones external to the utility specification, as in (1).

## 4.3  Estimating State-Dependence Parameters

### 4.3.1  Notation

For clarity, consumer and time subscripts are suppressed unless necessary. Denote coefficients,

subscripts and variables in the following manner:

| | |
|---:|:---|
| *Indices* | Time: $t = \{1, ..., T\}$, Brands: $j = \{1, ..., B\}$, Mix variables: $r = \{1, ..., R\}$ |
| *Variables* | Choice: $Y_{j,(t)} = 1$ if consumer buys brand $j$ at time $t$, 0 otherwise |
| | Marketing variables: $X_{rj}$ = the $r^{th}$ mix variable for brand $j$ |
| *Coefficients* | Marketing coefficients: $\{\beta_1, ..., \beta_R\}$, Brand dummies: $\{\beta_{01}, ..., \beta_{0B}\}$ |

### 4.3.2  Choice Probabilities

For the inertial model, the choice probabilities of each consumer are modeled as linear combinations of past choices and current choice probabilities with respect to all brands $i$ and time points $t$. Then,

the probability that household $k$ buys brand $i$ at time $t$ is given by $P_{it} = JY_{i(t-1)} + (1-J)\theta_{it}$, where:

$$\theta_{it} = \frac{\exp\left[\beta_{0i} + \beta_1 X_{1i} + ... + \beta_R X_{Ri}\right]}{\sum_{j=1}^{j=B} \exp\left[\beta_{0j} + \beta_1 X_{1j} + ... + \beta_R X_{Rj}\right]} \tag{20}$$

In this expression, we have suppressed the time subscripts on the marketing variables $X_{rj}$ with the understanding that these variables vary over time.

### 4.3.3 Optimization

For a single consumer, the log-likelihood function is given formally as follows, summing over brands and time:

$$\log(L) = \sum_{t=1}^{t=T} \sum_{j=1}^{j=B} Y_{j,t} \log(P_{j,t}) \tag{21}$$

A specific household's contribution to the overall gradient is therefore given as

$$\frac{\partial \log(L)}{\partial \beta_i} = (1-J) \sum_{t=1}^{t=T} \sum_{j=1}^{j=B} \frac{Y_{j,t}}{P_{j,t}} \theta_{jt} \frac{\partial \log(\theta_{jt})}{\partial \beta_i} \tag{22}$$

where:

$$
\begin{aligned}
\frac{\partial \log(\theta_{jt})}{\partial \beta_i} &= X_{ij} - \sum_{l=1}^{l=B} X_{il} \frac{\exp\left[\beta_{0l} + \beta_1 X_{1l} + ... + \beta_R X_{Rl}\right]}{\sum_{m=1}^{m=B} \exp\left[\beta_{0m} + \beta_1 X_{1m} + ... + \beta_R X_{Rm}\right]} \\
&= X_{ij} - \sum_{l=1}^{l=B} X_{il} \theta_{lt} \tag{23}
\end{aligned}
$$

The final expression in (23) represents the choice probability sensitivity to mix variables in a particularly attractive manner from the vantage point of estimation. It states implicitly that mix variables should be considered relative to what might be termed 'probabilistic centering', whereby the value of each is considered for the purposes of estimation as a deviation from that of other brands weighted by their (logit-based) choice probability. The $J$ parameter is modeled so as to incorporate an intrinsic bounding to the unit interval:

$$J = \frac{1}{1 + \exp(-\mu)} \tag{24}$$

Note that $\mu$ tending toward negative infinity indicates a value of $J$ near zero, and thus a pure zero-order logit model; conversely, $\mu$ tending toward positive infinity indicates a value of $J$ near one, and thus a pure first-order Markov model with no dependence on the marketing mix. It is necessary to calculate the following to complete the gradient:

$$\frac{\partial \log(L)}{\partial \mu} = \frac{\partial J}{\partial \mu} \sum_{t=1}^{t=T} \sum_{j=1}^{j=B} \frac{\partial}{\partial J} \log\left[JY_{i(t-1)} + (1-J)\theta_{it}\right] = J(1-J) \sum_{t=1}^{t=T} \sum_{j=1}^{j=B} \frac{Y_{j,t}}{P_{j,t}} \left[Y_{j(t-1)} - \theta_{it}\right] \tag{25}$$

To assemble the Hessian, second partials follow directly from these formulae though, for the purposes of optimization, it suffices, and is approximately as computationally-intensive, to assemble the component vectors of the Hessian by numerically differentiating (25) on a brand-by-brand basis.

## 4.4   Incorporating 'Non-Informative Priors'

In estimating the Inertial Logit model, an adjusted $\text{LOY}_{ik}$ measure must be used, to avoid logarithms of zero cell counts. It is possible to approach this in the manner of Fader and Lattin (1993), who estimate parameters to account for this, as well as nonstationarity, through the use of a Dirichlet-multinomial specification; this approach, despite the attractive feature of accounting for nonstationarity, is rather non-parsimonious, and we have taken a related path requiring fewer parametric estimates. Following a theory of non-informative priors developed by Box and Tiao (1973) in a Bayesian context, the loyalty measure takes a mildly altered form:

$$\text{LOY}_{ik} = \log \left[ \frac{\frac{1}{2} + \sum_t Y_{ikt}}{\frac{B}{2} + T_k} \right] \tag{26}$$

where $B$ is the number of brands. Although this measure may appear somewhat different from $\log(L_{ik})$, the probability limit of equation (26) is identical to that of equation (15); hence, the consistency-based results derived earlier remain unaffected by the alternative specification (26).

For comparison between models to be meaningful, care must be taken to ensure that all models have recourse to the same data. For example, the GL model makes predictions in a particular period as a function of all past periods, while the Simple and Inertial Logit models appeal to the entire choice history. To address this concern, we make use of both a calibration period and an estimation period, each consisting of 90 purchases. During the calibration period, both the simple-average and exponentially-smoothed (GL) loyalty variables are 'started up'. After the estimation period begins, only the GL loyalty variables are further updated, avoiding any potential confounds whereby a purchase at a given time is predicted through a measure calibrated using the outcome at that time; we note that such a scheme biases against the proposed loyalty method, and thus provides a conservative test of its efficacy.

## 4.5   Simulation Results

Table 1 summarizes the results for the 2 x 2 factorial design discussed previously. Under each condition, 50 independently-generated datasets were analyzed. Because the results were relatively

insensitive to changes in the marketing mix autocorrelation parameter $\rho$, we only present results for variation in the inertial parameter $J$.

| | J = .1 | | | J = .4 | | |
|---|---|---|---|---|---|---|
| | Inertial | Simple | GL | Inertial | Simple | GL |
| Intercept | −0.026 | −0.021 | −0.027 | 0.043 | 0.016 | −0.004 |
| std. error | 0.208 | 0.175 | 0.165 | 0.212 | 0.094 | 0.066 |
| Loyalty | 1.646 | 1.483 | 2.730 | 1.567 | 1.176 | 1.159 |
| std. error | 0.146 | 0.080 | 0.907 | 0.184 | 0.039 | 0.122 |
| Price | −2.038 | −1.673 | −1.649 | −1.974 | −0.892 | −1.022 |
| std. error | 0.187 | 0.120 | 0.116 | 0.240 | 0.115 | 0.119 |
| Carryover* | 0.092 | | 0.892 | 0.391 | | 0.061 |
| std. error | 0.030 | | 0.262 | 0.028 | | 0.069 |
| Log Likelihood | −418.83 | −432.59 | −433.15 | −430.51 | −522.85 | −459.76 |
| std. error | 19.62 | 17.76 | 16.83 | 17.38 | 15.74 | 17.28 |

**Table 1: Simulation Results**

* $J$ for Inertial Logit; $\lambda$ for GL; none for Simple Logit

These results provide compelling evidence that the proposed loyalty measure, relative to the Inertial Logit model structure, accurately recovers the parameters of the choice process. Evaluating the overall fit of the models in terms of log likelihood statistics, one finds that the Inertial Logit model fits best in both the $J = .1$ and $J = .4$ conditions. A comparison between the Inertial Logit and Simple Logit models using the likelihood ratio test finds that the Simple Logit fit is significantly poorer ($p < 10^{-6}$ for $J = .1$; $p \approx 0$ for $J = .4$). A comparison of the non-nested GL Logit and Inertial Logit models using the Rust and Schmittlein (1985) Bayesian Cross-Validated Likelihood (BCVL) test indicates that the Inertial Logit is the 'incorrect' model with $p < 10^{-6}$ for $J = .1$ and $p < 10^{-12}$ for $J = .4$.

Known values provide certain expectations regarding the values of the estimated parameters. In the Inertial Logit model, the intercepts represent the relative marketing effort of the brands. Because the price difference in the brands is set, on average, to zero, brand-specific intercept is expected to take this value as well; further, the known value of the Price coefficient is $-2.0$. Recovery of these parameters by the Inertial Logit model is quite good. In both the $J = .1$ and $J = .4$ conditions, the intercept is not statistically distinguishable from the correct value of zero. In terms of the price ($\beta$) and inertia ($J$) coefficients, the average values of the ($\beta, J$) estimates are ($-2.038, .092$) in the ($-2, .1$) condition, and ($-1.974, .391$) in the ($-2, .4$) condition, which again represent excellent

recovery. The one discrepant coefficient is associated with the loyalty variable itself; recall that theory implies that this coefficient be assigned a value of 1. The estimated values of 1.646 ($J = .1$) and 1.567 ($J = .4$) would at first seem to indicate a serious bias. However, this overestimate is an apparent artifact of Jensen's inequality, $E\left[g\left(Y\right)\right] \leq g\left[E\left(Y\right)\right]$, where $g\left(\cdot\right)$ is a concave function of the random variable $Y$. As applied here, $Y$ is the proportion of times the household purchased a particular brand, $g(Y) = \log(Y)$ is the loyalty measure used in the Inertial Logit model, and $g\left[E\left(Y\right)\right]$ is equal to the household's true brand preference. Jensen's inequality states that the expected value of the loyalty measure cannot be larger than the true brand preference parameter in any finite sample; it is important to note that this does not contradict the fact that the loyalty measure is a consistent estimate as the sample size becomes infinite. For this reason, the estimated value of the loyalty coefficient will tend to be greater than one to compensate for the finite sample bias in the loyalty measure. Hence, the observed overestimate is reasonable, and is in fact echoed in the empirical applications to follow.

## 4.6   Comparison with Benchmark Models

A comparison of the GL Logit with the Inertial Logit yields some degree of insight into the differences in the way habit-persistence (inertia) enters the probability specification in these two models. In keeping with the original parameterization in the GL Logit, larger values of $\lambda$ represent a lower weight on the last choice (relative to the long-run average over the entire choice history). That is, $\lambda = 0$ corresponds to all weight being placed on the previous choice, while $\lambda = 1$ corresponds to all previous choices being weighted equally. The GL model estimates $\lambda$ as 0.892 and 0.061, respectively, in the $J = .1$ and $J = .4$ conditions. This might at first appear to indicate that the parameter in the GL model acts as a proxy for $J$. However, a histogram of the individual runs (not shown) from which these average values are computed shows that the individual $\lambda$ values are clustered near the extremes of its allowable range (that is, near 0 or 1). As such, the GL model's loyalty variable seems to have been forced into an either-or situation, in which the model must 'choose' between a loyalty measure representing the average of the previous purchases and one that acts as a one-period lagged purchase indicator variable.

The parameter for which it is crucial to have accurate measurement is the Price coefficient, here the proxy for all marketing mix activity and, by implication, market-level sensitivity to managerial

control variables. While the Inertial Logit model reproduces the correct value of $\beta = -2$ quite accurately, this is not the case for the benchmark models. As one might expect, the Simple Logit model (which ignores choice-based feedback) underestimates the impact of Price ($b = -1.673$ ($J = .1$) and $b = -0.892$ ($J = .4$)). Surprisingly, the GL model, with its explicit mechanism for incorporating first-order choice dependence, does not fare much better. The fitted values of $b = -1.649$ ($J = .1$) and $b = -1.022$ ($J = .4$) indicate that the exponentially-smoothed loyalty variable formulation does not provide an adequate method for controlling for the choice feedback encoded in these data, although it must be borne in mind that this particularly parsimonious specification also attempts to account simultaneously for nonstationarity and preference heterogeneity.

## 4.7   Summary

The simulation experiment suggests two observations. First, as demonstrated in Table 1, the Inertial Logit loyalty procedure can accurately recover the key parameters of the choice process ($\beta$ and $J$) despite the upward bias on the loyalty variable coefficient. Second, the method works well in comparison to the selected benchmark models. The widely-applied loyalty measure first put forth by Guadagni and Little seems to perform in manner nearer to the Simple Logit than to the Inertial Logit – even though GL makes use of a type of first-order choice dependency. The key observation is that the GL loyalty measure is designed to account for both household preference heterogeneity and nonstationarity, while the Inertial Logit model framework only requires the loyalty measure to recover long-run preferences. The versatile role assumed by the GL loyalty measure may in fact account for its difficulties in recovering the choice parameters of the simulated data, despite its explicit encoding of first-order carryover effects.

# 5   Empirical Application

Although the simulated experiment provides evidence that the proposed loyalty measure leads to reasonably accurate recovery of choice parameters, it does not illustrate how the approach works with 'real world' choice data. In this section, the methodology developed thus far is applied to choice histories from a consumer panel. To allow for comparison to the simulation experiment of the previous section, the same models are estimated: Inertial Logit, Simple Logit and Guadagni-Little (GL) Logit.

## 5.1 Data Description

The data consist of household-level choices and store environment variables for two paper products categories, toilet tissue and paper towels. In each of these categories, there are eleven brands, both national brands and store brands, with wide variation in market share across brands. The data were obtained from a consumer panel in Ontario, Canada, covering a one-year period. Random samples of 160 households in the toilet tissue category and 182 households in the paper towels category were selected for analysis. Three marketing mix variables are available for use in estimation: price (cents per equivalent unit), in-store feature $(0-1)$ and in-store display $(0-1)$. Brand loyalties were computed from the observed brand purchases as noted below.

## 5.2 Computation of Loyalty Variables

Due to the relatively short purchase histories of the panel households, more explanation is required as to how the loyalty variables are calculated. For the Inertial Logit and Simple Logit models, household-level brand loyalties are taken to be the logarithm of the proportion of purchases allocated to each brand over the household's entire purchase record, modified for zero cell counts as in (26). Given the implications of the statistical theory developed thus far, there is no need to restrict the construction of these loyalty variables to an initial calibration period. A problem arises, however, in comparing the results of such a procedure with the analogous procedure suggested by Guadagni and Little (1983). The GL approach does in fact make use of a calibration period to 'start up' the loyalty variables; predictions of brand loyalty in a given period refer only to past periods. Since the loyalty variables in the Inertial and Simple Logit models require no calibration period, estimates of relative brand preference for early purchases will likely be considerably more accurate than the corresponding GL estimates. With this in mind, we have fit two variants of the GL model. The first assumes that, for the calculation of the loyalty variables, the share distribution at $t = 0$ is $1/B$, where $B$ is the number of brands; this, in fact, is the method used in the original GL model. It has the attractive feature of making no reference to external knowledge about the brands or households at $t = 0$. However, from a practical point of view, it should not be used unless a calibration dataset is available. We refer to this method as "GL Logit with $1/B$ Start-up." The second variant replaces the initial equal-share distribution vector at $t = 0$ with the average choice share for that household over the entire purchase record. The initial values for this version of the GL model make use of

the same data as that for the Inertial and Simple Logit models. The key difference is that the GL loyalty measure is updated over time as new purchases are recorded. We refer to this method as the "GL Logit with HH Start-up." Although both GL variants are shown in the results, we believe that only this second variant can be meaningfully compared to the Inertial and Simple Logit models in the analysis to follow.

## 5.3   Model Comparison

Table 2 lists parameter estimates and standard errors for the four models applied to the paper towel and toilet tissue categories; for clarity, we have omitted the 10 brand-specific intercepts that were estimated for each of the models. Notice that the fit of the "GL with $1/B$ Start-up" model (in terms of log likelihood) is far poorer than the others. For reasons noted above, we hold aside this GL Logit variant in the following discussion.

| Table 2: Parameter Estimates for Paper Goods Choice Histories | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Paper Towels** | | | | **Toilet Tissue** | | | |
| | Inertial | Simple | GL-$1/B$ | GL-HH | Inertial | Simple | GL-$1/B$ | GL-HH |
| Loyalty | 1.909 | 1.819 | 4.040 | 5.626 | 1.669 | 1.543 | 3.866 | 5.325 |
| std. error | 0.086 | 0.077 | 0.290 | 0.224 | 0.049 | 0.047 | 0.180 | 0.162 |
| Price | -0.983 | -0.919 | -1.134 | -0.944 | -0.524 | -0.532 | -0.579 | -0.650 |
| std. error | 0.162 | 0.178 | 0.173 | 0.193 | 0.097 | 0.082 | 0.079 | 0.088 |
| Feature | 0.946 | 0.728 | 0.600 | 0.563 | 1.859 | 1.678 | 1.588 | 1.530 |
| std. error | 0.186 | 0.186 | 0.152 | 0.193 | 0.171 | 0.151 | 0.130 | 0.149 |
| Display | 0.590 | 0.659 | 0.741 | 0.637 | 0.846 | 0.781 | 0.697 | 0.814 |
| std. error | 0.197 | 0.183 | 0.153 | 0.191 | 0.174 | 0.157 | 0.136 | 0.155 |
| Carryover[a] | 0.037 | | 0.766 | 0.999 | 0.039 | | 0.758 | 0.999 |
| std. error | 0.016 | | 0.026 | —[b] | 0.014 | | 0.019 | —[b] |
| Log Likelihood | -643.6 | -716.8 | -1146.8 | -697.3 | -1511.5 | -1559.5 | -2163.4 | -1604.3 |
| [a]Carryover parameter is $J$ for Inertial Logit and $\lambda$ for GL; Carryover does not enter into the Simple Logit. | | | | | | | | |
| [b]Parameter was estimated to lie at the boundary of its allowable range. | | | | | | | | |

There are two principal conclusions to be discerned from Table 2. The first of these involves a comparison of log-likelihood values across the models.[2] Here the story is unequivocal: the Inertial Logit model fares far better than the Simple Logit and GL Logit models. In the paper towels data, both the Inertial and GL models fit significantly better than the Simple Logit model ($p \approx 0$ and

---

[2]The differences in log-likelihood between the various models are large enough, relative to differences in number of parameters, that AIC and BIC measures yield identical conclusions, and so have been omitted.

$p < 10^{-9}$, respectively). The Inertial Logit and GL models can be compared using the BCVL procedure (Rust and Schmittlein, 1985); the probability of the Inertial Logit model being 'incorrect' is less than $10^{-23}$, and a similar conclusion holds for the toilet tissue data. A likelihood ratio test comparing the Inertial Logit to the Simple Logit provides strong evidence for choice feedback ($p \approx 0$). Moreover, the BCVL procedure favors the Inertial Logit model over the GL Logit with probability near 1.

It is also instructive to compare these models in terms of the role of the first-order dependence, represented by $J$ (Inertial Logit) and $\lambda$ (GL Logit). For both data sets, the inertial parameter $J$ assumed modest values (approximately .04). However, comparing the Simple Logit to the Inertial Logit, it is clear that these modest values of $J$ differ significantly from zero; in fact, the likelihood function peaks sharply at these values. For the "GL Logit with HH Start-up" models, an endpoint ($\lambda \approx 1$) was reached when the initial loyalty values were set to the household-level means (i.e., all weight was placed on the mean and none on the previous purchase). Put differently, when a 'meaningful' loyalty measure is used at $t = 0$, holding loyalty at the long-run average of observed choice yields the best fit for these data. Thus, for both datasets, the GL Logit approximates the Simple Logit model with no choice inertia, while the Inertial Logit model indicates a modest degree of first-order carryover.

## 5.4 Prediction of Inertial Logit Intercepts

Although the Inertial Logit model fits the data well relative to the benchmark models, this fact does not directly support the loyalty variable theory developed previously. However, the estimated brand intercepts of the Inertial Logit provide some evidence that the model is not overtly misspecified. Recall that the intercepts are estimates of $\tau_i = -\log(M_i)$, where $M_i = E\left[\exp\left(\mu_{it}\right)\right]$ depends upon the random stationary component of the Brand $i$ marketing mix effort variable $\sum_r \beta_r X_{rit}$. Using the statistical theory of cumulant generating functions (e.g., Kendall and Stuart 1977), it can be shown that, to first-order:

$$\log(M_i) \sim E\left[\mu_{it}\right] \tag{27}$$

To this order of approximation, we can apply equation (11) to write:

$$\tau_i \sim \varphi - \sum_r \beta_r E\left[X_{rit}\right] \tag{28}$$

23

where $\varphi = E[\phi_t]$ does not depend upon brand. That is, the brand intercepts should be negatively correlated with the expected value of the marketing mix effort variables $\sum_r \beta_r E[X_{rit}]$.

To examine this prediction, it is possible to compare the estimated intercepts from the Inertial Logit model to the expected effort variable $\sum_r \beta_r E[X_{rit}]$. In computing the marketing effort variable, one can make use of the marketing mix coefficient estimates from the Inertial Logit model ($\beta_r$) and the brand-by-brand sample temporal means $E[X_{rit}]$ of the marketing mix variables (i.e., price, feature and display). The correlations between the estimated model intercepts and the computed effort variable $\sum_r \beta_r E[X_{rit}]$ are $-.68$ (toilet tissue) and $-.74$ (paper towels). Both correlations are significantly negative, based upon standard statistical tests. Thus, the estimated intercepts do appear to be corrections for the differential marketing activity of the brands.

## 5.5   Summary

Clearly, we cannot know the true process generating household-level choice histories in these paper goods categories. However, the results of this application are encouraging in this regard. The Inertial Logit model fits well relative to benchmark models, and its intercepts are related to brand marketing mix activity as predicted by the developed theory.

## 6   Conclusions

The present research develops a general approach to the construction of loyalty variables to represent household-level heterogeneity in choice models. Conceptually, a logit loyalty measure is defined as any function of observed choices which provides a consistent estimate of the household's brand-specific intercepts. Drawing upon an Inertial Logit model framework (e.g., Roy, Chintagunta and Haldar 1996), we derive a loyalty variable procedure and examine its statistical properties using both simulated and actual choice histories.

A major implication of this work is that loyalty variables are difficult to construct on *a priori* grounds, and that such constructions may not be practicable for certain types of choice processes. In the Inertial Logit model, the derived loyalty measure incorporates the effects of two factors: intrinsic household-level preferences, and the impact of marketing mix activity. Because this decomposition can be linearized and made independent of the degree of choice inertia, a procedure can be developed which leads to unbiased estimates of the parameters of the choice process. For an arbitrarily-

specified choice process, neither of these properties need hold. A strict reading of these results is that loyalty variables cannot be advocated independently of the choice process from which they are derived; hence, a compelling reason to employ more sophisticated heterogeneity adjustments (e.g., parametric or semi-parametric heterogeneity distributions) is the lack of an explicit loyalty variable theory for the choice process under consideration. This work also serves as a caveat to researchers interested in decomposing logit model brand intercepts into product attribute utilities. Such models have been suggested for applications in brand equity measurement (Kamakura and Russell 1993) and new product forecasting (Fader and Hardie 1996), as well as dozens of related applications since. Both the theoretical discussion and empirical results presented support the notion that global brand intercepts, such as the $\tau_i$ in the Inertial Logit model, can be conceptualized solely as functions of marketing mix activity when loyalty variables are present in the model. For this reason, researchers interested in 'conjoint-like' models for scanner panel data should employ a degree of caution when loyalty variables are included in the model specification. For these types of problems, a random coefficient representation of heterogeneity may be more appropriate than the loyalty variable approach developed here.

The presented empirical findings should not be taken to indicate that the loyalty variable specification put forth by Guadagni and Little is an inappropriate choice for scanner panel data analysis. On the contrary, in a world of nonstationary brand preferences or pronounced new product activity, an exponentially-smoothed measure of household preference may well be a superior choice in building a flexible and parsimonious model. In the present context, the performance of the GL Logit serves mainly to illustrate the comparative properties of the Inertial Logit model. The improved fit of the Inertial Logit relative to the GL Logit benchmark may in fact be a manifestation of the Inertial Logit's separation of two distinct constructs: brand preference and habit-persistence. Intuitively, this explains why the Inertial Logit specification allows the construction of a loyalty measure which is independent of the degree of inertia in the choice process. The Inertial Logit model specification may itself prove useful in future research, and the present work provides a readily-applicable methodology for estimating such models, one that can in fact can be extended to a generalized class of logit-like models in which an inertial component can be parsimoniously grafted onto a variety of utility specifications.

Appendix: Unconditional Expectation of Purchase Indicator Variable

Here we wish to show that the unconditional expectation of the household's purchase indicator vector is independent of the inertial parameter $J$. Define the logit choice probability vector and the observed purchase vector of household $k$, respectively, as:

$$\tilde{\theta}_{kt} = [\theta_{1kt}\ \theta_{2kt}\ ...\ \theta_{Bkt}]$$

$$\tilde{Y}_{kt} = [Y_{1kt}\ Y_{2kt}\ ...\ Y_{Bkt}]$$

Using this notation, the Inertial Logit model can be described succinctly as:

$$\tilde{Y}_{kt} \quad | \quad \tilde{Y}_{k(t-1)}, \tilde{\theta}_{kt} \sim Multinomial\,(1, P_{kt})$$

$$P_{kt} = JY_{k(t-1)} + (1 - J)\theta_{kt}\ ;\ \ 0 \leq J \leq 1$$

We assume that the expectation of the logit choice probability vector $\tilde{\theta}_{kt}$ is independent of time. In symbols, $E\left[\tilde{\theta}_{kt}\right] = \tilde{\theta}_k$. Using this assumption and the fact that $E\left[\tilde{Y}_{kt} \mid \tilde{Y}_{k(t-1)}, \tilde{\theta}_{kt}\right] = P_{kt}$, we obtain the expression:

$$E\left[\tilde{Y}_{kt}\right] = JE\left[\tilde{Y}_{k(t-1)}\right] + (1 - J)\,\tilde{\theta}_k$$

This is a recursion relation in the successive expectations. It is equivalent to

$$E\left[\tilde{Y}_{kt}\right] = J^n E\left[\tilde{Y}_{k(t-n)}\right] + (1 - J^n)\,\tilde{\theta}_k$$

Since the expectations are bounded between zero and one, allowing $n$ to become arbitrarily large demonstrates that the unconditional expectation is independent of the inertial parameter $J$:

$$E\left[\tilde{Y}_{kt}\right] = \tilde{\theta}_k$$

# References

[1] Abe, Makoto, Yasemin Boztug and Lutz Hildebrandt (2003), "Investigating the Competitive Assumption of Multinomial Logit Models of Brand Choice by Nonparametric Modeling," working paper, University of Tokyo, Faculty of Economics.

[2] Abramson, Charles, Rick L. Andrews, Imran S. Currim, and Morgan Jones (2000), "Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice," Journal of Marketing Research, vol. 37, no. 4, Nov., p. 410-426.

[3] Allenby, Greg M. and Peter J. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89 (December), 1218-1231.

[4] Allenby, Greg M. and Peter E. Rossi (1991), "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science*, 10 (Summer), 185-204.

[5] Anderson, Eric (2002), "A Guadagni-Little Likelihood Can Have Multiple Maxima," working paper, Graduate School of Business, University of Chicago.

[6] Bell, David and James Lattin (2000), "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity," Marketing Science, vol. 19, no. 2, Spring, pp. 185-200.

[7] Box, George P. and George C. Tiao (1973), Bayesian Inference in Statistical Analysis, Reading, Mass.: Addison-Wesley Publishing Co.

[8] 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.

[9] Carpenter, Greg and Donald Lehman (1985), "A Model of Marketing Mix, Brand Switching and Competition," *Journal of Marketing Research*, 22, 318-329.

[10] Fader, Peter and Bruce Hardie (1996), "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33(4), 442-452.

[11] Fader, Peter and James Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Science*, 12 (Summer), 304-317.

[12] Fader, Peter, James M. Lattin and John D. C. Little (1992), "Estimating Nonlinear Parameters in the Multinomial Logit Model," *Marketing Science*, 11 (Fall), 372-385.

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

[14] Guadagni, Peter and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203-238.

[15] Heckman, J. J. (1981), "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Applications*, C. F. Manski and D. McFadden (eds.), Cambridge: MIT Press, 114-178.

[16] Jeuland, Abel (1979), "Brand Inertia as One Aspect of the Notion of Brand Loyalty," *Management Science*, 25 (July), 671-682.

[17] Jones, Morgan and J. T. Landwehr (1988), "Removing Heterogeneity Bias from Logit Model Estimation," *Marketing Science*, 7 (Winter), 41-59.

[18] Kamakura, Wagner A. and Gary J. Russell (1993), "Measuring Brand Value With Scanner Data," *International Journal of Research in Marketing*, 10 (March), 9-22.

[19] Kamakura, Wagner and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379-390.

[20] Kendall, Maurice and Alan Stuart (1977), The Advanced Theory of Statistics, Volume 1, New York: MacMillan Publishing Co.

[21] Kim, Jin Gyo, Ulrich Menzefricke and Fred M. Feinberg (2003), "Capturing Choice Dynamics Through a Bayesian Logit Model with Time-Varying Parameters," working paper, MIT Sloan School of Management.

[22] Lattin, James and Randolph Bucklin (1989), "Reference Effects of Price and Promotion on Brand Choice Behavior," *Journal of Marketing Research*, 26 (August), 299-310.

[23] Rossi, Peter E. and Greg M. Allenby (2003) "Bayesian Statistics and Marketing," *Marketing Science*, forthcoming.

[24] Roy, R., Chintagunta, P.K., Haldar, S. (1996), "A Framework for Investigating Habits, "The Hand of the Past," and Heterogeneity in Dynamic Brand Choice," *Marketing Science*, 15(3), 280-299.

[25] Rust, Roland and David C. Schmittlein (1985),"A Bayesian Cross-Validated Likelihood Method for Comparing Alternative Specifications of Quantitative Models," *Marketing Science*, 4 (Winter), 20-40.

[26] Seetharaman, P.B. (2003), "Modeling Multiple Sources Of State Dependence In Random Utility Models: A Distributed Lag Approach," *Marketing Science*, forthcoming.

[27] Tellis, Gerard J. (1988),"Advertising Exposure, Loyalty and Brand Purchase: A Two-Stage Model of Choice," *Journal of Marketing Research*, 25 (May), 134-144.