The Relative Effects of State Dependence and Habit Persistence on Mean Convergence in First-Order Models of Brand Choice

Fred M. Feinberg

University of Michigan Business School
701 Tappan Street, E2422
Ann Arbor, MI 48109
feinf@umich.edu

The author wishes to thank Gary Russell and Pete Fader for helpful comments on earlier versions of this manuscript. The usual disclaimer applies.

The Relative Effects of State Dependence and Habit Persistence on Mean Convergence in First-Order Models of Brand Choice

Abstract

The present paper examines two methods for encoding first-order dependence into models of brand choice: Habit Persistence, based on temporal autocorrelation in the probability vector predicted by a baseline model; and State Dependence, based on correlation between such predicted choices and the vector of previously chosen brands. As a measure of convergence rate, the variance of the mean estimator in the binary choice case is calculated in closed form, demonstrating that the convergence rate-dampening effects of state dependence are greater than those of habit persistence; further, the two first-order carry-over methods operate synergistically, though differently, in various regions of their joint parameter space.

1 Introduction

In a recent paper, Roy, Chintagunta and Haldar (1996) put forth a framework for the analysis of scanner panel data that relies on three separate constructs, Habit Persistence, State Dependence and household-level Heterogeneity, demonstrating that the specifications advocated are in fact those of utility-maximization. Methods to account for heterogeneity have been a major research topic in Marketing over the past decade, beginning with Guadagni and Little's (1983) account of 'loyalty' variables and culminating in a number of sophisticated statistical techniques of varying degrees of parameterization; full treatments of this literature can be found in Chintagunta, Jain and Vilcassim (1991), Fader and Lattin (1993), Gonul and Srinivasan (1993), and Allenby and Lenk (1994). However, a good deal less attention has been focused on the remaining constructs, habit persistence and state dependence.

The present paper addresses the mean vector of household-level purchase indicator variables, a measure important not only as a zero-order estimate of disaggregate choice shares, but in the calculation of heterogeneity corrections descending from Guadagni and Little's original loyalty specification. For example, Feinberg and Russell (1997) have formulated a model which takes such uncorrected (by marketing mix activity) choice shares as a proxy for models of the latent class type (e.g., Kamakura and Russell 1989; Lenk and DeSarbo 1996), which largely dispense with direct household-level measures based on past purchases. Among the concerns in utilizing such variables are their degree of bias and rate of convergence, given the small number of per-household observations typical in consumer panels. Positing a choice model formulation based on that explored by Jeuland (1979), Roy et. al. (1996) and Feinberg and Russell (1997), explicit measures of mean bias and variance, as functions of the degree of habit persistence and level of state dependence, can be obtained.

2 Modeling Framework

Recall that, in the standard framework under which methods such as Logit or Probit are applied to scanner panel choice data, the underlying choice process is taken to be multinomial. That is, each household is presumed, on a particular choice occasion, to choose from the available brands in a manner which maximizes (stochastic) utility; the deterministic component is typically a function, linear-in-coefficients, of observable marketing mix activity variables, a set of brand intercepts measuring intrinsic market-level preferences and, depending on the particular utility specification, a method of encoding household-level heterogeneity, for instance a set of possible 'purchase indicator' variables calculated from the household's past choices. Standard one-period (or greater) lag formulations and loyalty variables are examples of this last set of utility components. Apart from such a utility specification, choice between models amounts to the various methods of encoding error which, for example, can be taken as double exponential or multinormal. The brand choice process can then be taken as multinomial with parameter $\tilde{\theta}_t$, where the time (t) subscript is suppressed where understood.

We consider, for purposes of exposition, choice between two relevant brands; extension to multiple brands can be considered as a nested set of choices between a focal brand and all others, consistent with, for example, Luce-type scaling or the IIA property. As developed in Resnick and Roy (1991), an 'inertial' component in the choice process, representing state dependence, can be accommodated peripherally, with household-level choice probabilities expressed as a linear function of that given by, for example, the standard Logit and a one-period lagged purchase indicator variable, Y_{t-1}.

2.1 Specification of Habit Persistence and State Dependence

It is important to differentiate between two fundamentally different types of carry-over arising in household-level purchase modeling, that of habit persistence intrinsic to the underlying multinomial choice process generated by the utility specification (e.g., normed attractions), and that arising from a state-dependent component that is, in some sense, overlaid upon the multinomial process. Habit persistence, which can arise from temporal correlations in predictor variables and thus enter into the utility specification directly (although this need not be the case), can be thought of as endogenous. By contrast, state dependence, as formulated as 'inertia' by Jeuland (1979), in a more general setting by Heckman (1981), and extended to a generalized (logistic) brand choice framework by Roy, Chintagunta and Haldar (1996), is a temporal property of the household-level multinomial choice vector itself, in terms of its correlation with the previous choice (for additional detail in the context of variety-seeking and inertial models, see Hutchinson 1986, or Bawa 1990). Here, we explore the theory without reference to a particular error specification, and so consider habit persistence in a form that dispenses with the particularities of the predictor variables themselves, specifically in how they are related functionally to the multinomial process generating $\tilde{\theta}$.

With the degree of state dependence in the process represented by J and that of habit persistence by ρ, choice on a particular occasion between two relevant brands, Y_t, is binomial:

$$Y_t \sim B\left[1, (1-J)\theta_t + JY_{t-1}\right]$$

$$\theta_t = (1-\rho)P_t + \rho\theta_{t-1} \qquad for\ t = 1, ..., n \tag{1}$$

$$P_t \sim f_P(p)\ i.i.d. \qquad \theta_0 = P_0 = E[P] \equiv p_0 \qquad Var[P_t] = \sigma_p^2$$

There is an underlying generating process, given by the $\{P_t\}$, each with time-invariant density $f_P(p)$; the $\{\theta_t\}$ are merely geometrically-weighted sums of independent draws from $f_P(p)$, and are thus autocorrelated. Analogous to the process put forth by Jeuland (1979), the parameter J allows the generating multinomial model to be nested (i.e., $J = 0$), as well as for incorporating 'perfect' state dependence (i.e., $J = 1$, whence the Y_t are identically zero or one). Toward the end of establishing the degree of bias and rate of convergence of the household-level mean estimator, $\overline{Y}$, it is necessary to calculate its expected value and standard error. Considering the recursion relationships (1), it is easily verified that the following hold:

$$\theta_t = \rho^t P_0 + (1-\rho)\sum_{i=1}^{i=t} \rho^{t-i} P_i$$

$$E\left[\theta_t\right] = p_0 \qquad Var[\theta_t] = \frac{1-\rho}{1+\rho}\left(1-\rho^{2t}\right)\sigma_p^2 \rightarrow \frac{1-\rho}{1+\rho}\sigma_p^2 \tag{2}$$

$$Cov[\theta_t, \theta_{t+k}] = \rho^k \frac{1-\rho}{1+\rho}\left(1-\rho^{2t}\right)\sigma_p^2 \rightarrow \frac{1-\rho}{1+\rho}\rho^k\sigma_p^2$$

Note that, in the event that there is no explicit habit persistence, $\rho = 0$, and the $\{\theta_i\}$ are identical to the $\{P_i\}$ series and inherit their independence properties. This simpler case is treated first. Before doing so, it is useful to state the following fact about conditional expectations for recursion relations arising from the process given by (1) above. The conditional mean vector of the purchase indicator variable $[Y_t \mid Y_{t-1}]$ can be expressed in matrix notation as follows:

$$E\left(\left[\begin{array}{c} 1-Y_t \\ Y_t \end{array}\right] \mid Y_{t-1}\right) = \left[\begin{array}{cc} J + (1-J)(1-\theta_t) & (1-J)(1-\theta_t) \\ (1-J)\theta_t & J + (1-J)\theta_t \end{array}\right]\left[\begin{array}{c} 1-Y_{t-1} \\ Y_{t-1} \end{array}\right] \tag{3}$$

Referring to the transition matrix in (3) as A_t, a higher-order conditional relationship can be written:

$$E\begin{bmatrix} 1-Y_{t+k} \\ Y_{t+k} \end{bmatrix} | Y_t = [A_{t+k}A_{t+k-1} \ldots A_{t+1}]\begin{bmatrix} 1-Y_t \\ Y_t \end{bmatrix} \tag{4}$$

The problem is substantially simplified by noticing the following special form, easily verified by induction, where the off-diagonal elements are chosen so that the columns sum to one:

$$\begin{aligned} & [A_{t+k}A_{t+k-1} \ldots A_{t+1}] \\ = & \begin{bmatrix} J^k + (1-J)\left[(1-\theta_{t+k}) + J(1-\theta_{t+k-1}) + \ldots + J^{k-1}(1-\theta_{t+1})\right] & * \\ * & J^k + (1-J)\left[\theta_{t+k} + J\theta_{t+k-1} + \ldots + J^{k-1}\theta_{t+1}\right] \end{bmatrix} \end{aligned} \tag{5}$$

3 Special Case: State Dependence Alone

When $\rho = 0$, the $\{\theta_i\}$ are independent draws from $f_P(p)$, so that the expected value of a product of the matrices expressed in (5) is merely the product of their expected values, each of which is a constant Markov matrix with p_0 in place of any θ_i. Thus, the covariance of any particular pair of purchase indicator variables can be shown to be given by:

$$Cov\left[Y_t, Y_{t+k}\right] = J^k p_0(1-p_0)(1-J^{2t}) \longrightarrow J^k p_0(1-p_0) \tag{6}$$

To compute the asymptotic variance (w.r.t. t) of the mean estimator, it is useful to appeal to a sum of covariances:

$$Var\left[\overline{Y}\right] = \frac{1}{n^2}\sum_{t<i,j\leq t+n} Cov\left[Y_i, Y_j\right] = \frac{p_0(1-p_0)}{n^2}\sum_{t<i,j\leq t+n} J^{|i-j|} \tag{7}$$

The values of the summand in the double summation (7) form a symmetric Toeplitz matrix generated by the row vector $[1, J, \ldots, J_{n-1}]$ (holding aside t, which can be taken to be arbitrarily large), where the (i,j) entry is given by $J^{|i-j|}$; for example, for $n = 4$:

$$\begin{bmatrix} 1 & J & J^2 & J^3 \\ J & 1 & J & J^2 \\ J^2 & J & 1 & J \\ J^3 & J^2 & J & 1 \end{bmatrix} \tag{8}$$

Computing the summation in (7) results in an expression for the standard error:

$$Var\left[\overline{Y}\right] = \frac{p_0(1-p_0)}{n^2}\left[\frac{n(1-J^2) - 2J(1-J^n)}{(1-J)^2}\right] \tag{9}$$

The expression (9), curiously enough, varies with n to an order that is dependent on the relative values of J and of n itself. For small values of J or large values of n, (9) is locally approximated by:

$$Var\left[\overline{Y}\right] = \frac{p_0(1-p_0)}{n}\left(\frac{1+J}{1-J}\right) \tag{10}$$

For values of J close to 1 and small values of n, (9) is, to first order:

$$Var\left[\overline{Y}\right] = p_0(1-p_0)\left[1-(1-J)(2n-1)\right] \tag{11}$$

An intriguing property of (10 - 11) concerns the rate of convergence of the mean estimator (as measured by its standard error) with respect to the sample size n which, while on the order of $n^{-1/2}$ for small quantities of the state dependence parameter and large sample sizes, is essentially invariant with n when state dependence is pronounced and sample sizes are relatively small. Thus, empirical estimates of the degree of state dependence typical in panel data studies becomes an important determinant of the 'quality' of the household-level mean estimator. Although we do not discuss methods of estimating the degree of state dependence or habit persistence here, these issues are addressed by both Roy, Chintagunta and Haldar (1996) and Feinberg and Russell (1997).

4 Full Model: State Dependence and Habit Persistence

Notice that, when $\rho \neq 0$, the θ_i are no longer independent draws from $f_P(p)$, but follow the recursion relationships given by (1); thus, the expected value of the product of the transition matrices can no longer be expanded to a product of expected values, so that another route must be taken toward computing the covariance terms in the following sum:

$$Var\left[\overline{Y}\right] = \frac{1}{n^2}\sum_{t<i,j<t+n} Cov\left[Y_i, Y_j\right] \tag{12}$$

It is helpful to calculate covariance from its definition:

$$\begin{aligned} Cov\left[Y_t, Y_{t+k}\right] &= E\ \left[Y_t Y_{t+k}\right] - E\left[Y_t\right]E\left[Y_{t+k}\right] \ = \ E\left[Y_t Y_{t+k}\right] - p_0^2 \\ &= P\left[Y_t = Y_{t+k} = 1\right] - p_0^2 \end{aligned} \tag{13}$$

For this last probability to be computed, it is possible to condition on the joint distribution of the $\{\theta_i\}$ series (let $\tilde{\theta} = (\theta_0, ..., \theta_t, ..., \theta_{t+k})$):

$$\begin{aligned}
& P\left[Y_t = Y_{t+k} = 1\right] \\
= \; & E_{\tilde{\theta}}\left[P\left[Y_{t+k} = 1 \mid Y_t = 1; \tilde{\theta}\right] P\left[Y_t = 1 \mid \tilde{\theta}\right]\right] \\
= \; & E_{\tilde{\theta}}\left[\begin{array}{c} \left[J^k + (1-J)\left(\theta_{t+k} + J\theta_{t+k-1} + ... + J^{k-1}\theta_{t+1}\right)\right] \cdot \\ \left[J^t + (1-J)\left(\theta_t + J\theta_{t-1} + ... + J^{t-1}\theta_1\right)\right] \end{array}\right] \\
= \; & J^{t+k} + J^t(1-J)E_{\tilde{\theta}}\left[\theta_{t+k} + ... + J^{k-1}\theta_{t+1}\right] + J^k(1-J)E_{\tilde{\theta}}\left[\theta_t + ... + J^{t-1}\theta_1\right] + ... \\
& +(1-J)^2 E_{\tilde{\theta}}\left[\left(\theta_{t+k} + ... + J^{k-1}\theta_{t+1}\right)\left(\theta_t + ... + J^{t-1}\theta_1\right)\right]
\end{aligned} \tag{14}$$

All but the last of the four summed expressions above are readily evaluated using (2) as:

$$J^{t+k} + (1-J)\left[J^t\frac{1-J^k}{1-J} + J^k\frac{1-J^t}{1-J}\right]p_0 \longrightarrow J^k p_0 \tag{15}$$

All that remains is the calculation of the fourth component term in (14), the expectation of the product of the ρ-weighted future and past choice probabilities:

$$\begin{aligned}
& E_{\tilde{\theta}}\left[\left(\theta_{t+k} + ... + J^{k-1}\theta_{t+1}\right)\left(\theta_t + ... + J^{t-1}\theta_1\right)\right] \\
= \; & Cov\left[\left(\theta_{t+k} + ... + J^{k-1}\theta_{t+1}\right), \left(\theta_t + ... + J^{t-1}\theta_1\right)\right] \\
& +E_{\tilde{\theta}}\left[\theta_{t+k} + ... + J^{k-1}\theta_{t+1}\right] E_{\tilde{\theta}}\left[\theta_t + ... + J^{t-1}\theta_1\right] \\
= \; & \sum_{i=0}^{i=k-1}\sum_{j=0}^{j=t-1} J^{(i+j)} Cov\left[\theta_{t+k-i}, \theta_{t-j}\right] + \left[\sum_{i=0}^{i=k-1} p_0 J^i\right]\left[\sum_{j=0}^{j=t-1} p_0 J^j\right] \\
= \; & \left[\sum_{i=0}^{i=k-1}\sum_{j=0}^{j=t-1} J^{(i+j)}\rho^{(k-i+j)}\frac{1-\rho}{1+\rho}\left(1-\rho^{2(t-j)}\right)\sigma_p^2\right] + \frac{\left(1-J^k\right)\left(1-J^t\right)}{\left(1-J\right)^2}p_0^2
\end{aligned} \tag{16}$$

This double summation can be calculated by well-known formulas; upon letting t increase without bound, the resulting value for (16) is:

$$\frac{\rho\left(1-\rho\right)}{\left(\rho-J\right)\left(1-\rho J\right)\left(1+\rho\right)}\left(\rho^k - J^k\right)\sigma_p^2 + \frac{\left(1-J^k\right)}{\left(1-J\right)^2}p_0^2 \tag{17}$$

Thus, from (13) and (17), the required covariances are then given by:

$$\begin{aligned}
\lim_{t\longrightarrow\infty} Cov\left[Y_t, Y_{t+k}\right] &= \lim_{t\longrightarrow\infty} P\left[Y_t = Y_{t+k} = 1\right] - p_0^2 = \\
&= \frac{\rho\left(1-\rho\right)\left(1-J\right)^2}{\left(\rho-J\right)\left(1-\rho J\right)\left(1+\rho\right)}\left(\rho^k - J^k\right)\sigma_p^2 \quad + \quad J^k p_0\left(1-p_0\right)
\end{aligned} \tag{18}$$

Note that, when $\rho = 0$ or $\rho = 1$, (18) agrees with the corresponding expression given in (6). It is now possible to calculate the standard error of the mean estimator for positive values of ρ, in the same manner as when $\rho = 0$, by referring to the Toeplitz matrix (8) and the associated summation, (9). It therefore follows directly from (9) that:

$$\lim_{t \longrightarrow \infty} Var\left[\overline{Y}\right] = \left[\frac{n(1-J^2) - 2J(1-J^n)}{(1-J)^2}\right] \frac{p_0(1-p_0)}{n^2} \tag{19}$$

$$+\frac{\rho\,(1-\rho)\,(1-J)^2}{(\rho-J)\,(1-\rho J)\,(1+\rho)} \left[\frac{n(1-\rho^2) - 2\rho(1-\rho^n)}{(1-\rho)^2} - \frac{n(1-J^2) - 2J(1-J^n)}{(1-J)^2}\right] \frac{\sigma_p^2}{n^2}$$

Thus, the variance of the mean estimator is a weighted sum of variances, one representing the binomial process with P set at its expected value, $p_0(1-p_0)$, the other representing the distribution from which the P's are drawn, σ_p^2. The greatest deviation from the actual variance, by taking $\rho = 0$, will come about when the latter variance is as large as possible compared with the former one; this occurs when the (assumed-continuous) P distribution approaches degeneracy (in the form of a binomial, for instance, if it were Beta distributed), with variance bounded from above by $p_0(1-p_0)$. For purposes of illustration, therefore, we will consider this extreme case where $\sigma_p^2 = p_0(1-p_0)$, so that the degree to which the variance is underestimated by setting $\rho = 0$ can be assessed by computing the ratio of the weighting factors in (19). Define f, the proportional underestimation of the variance by taking ρ to be zero, as follows[1]:

$$f(J,\rho,n) = \frac{\rho\,(1-\rho)\,(1-J)^2}{(\rho-J)\,(1-\rho J)\,(1+\rho)} \left(\frac{n(1-\rho^2) - 2\rho(1-\rho^n)}{n(1-J^2) - 2J(1-J^n)} \frac{(1-J)^2}{(1-\rho)^2} - 1\right) \tag{20}$$

Note that f is non-negative[2]. Values of f near 0 indicate that f is well-approximated by $f(J,0,n)$, while "large" values of f indicate the opposite. The following can be established algebraically: (1) $f_J \leq 0$; (2) $f_n \geq 0$; (3) $f_\rho \geq 0$ for $\rho \approx 0$ and ≤ 0 for $\rho \approx 1$; (4) $f \leq 1$. Because $f_n \geq 0$, we have $f(J,\rho,2) \leq f(J,\rho,n) \leq \lim_{n\to\infty} f(J,\rho,n)$, so the following relations essentially "frame" f by

[1] Because $f(J,1,n) = f(J,0,n) = 0$, assuming a baseline case of $\rho = 1$ is identical to that of $\rho = 0$.

[2] When J and ρ are nearly equal, f is, to first order

$$\frac{2J\,(1-J)}{(1+J)^2}\left(1 - (1-J)\,\frac{nJ\left(1-J^{n-1}\right) + (1-J^n)}{n\,(1-J^2) - 2\,J\,(1-J^n)}\right)$$

which approaches $\frac{2J(1-J)}{(1+J)^3}$ for large n.

providing bounds for it:

$$\begin{aligned} f(J,\rho,2) &= \frac{\rho\,(1-\rho)\,(1-J)^2}{(1+J)\,(1-\rho J)\,(1+\rho)} \\ f(J,\rho,\infty) &= \frac{2\rho\,(1-J)^2}{(1+J)\,(1-\rho J)\,(1+\rho)} \end{aligned} \tag{21}$$

These functions are depicted, along with a representative plot of f (for $n = 10$), in Figure [1].

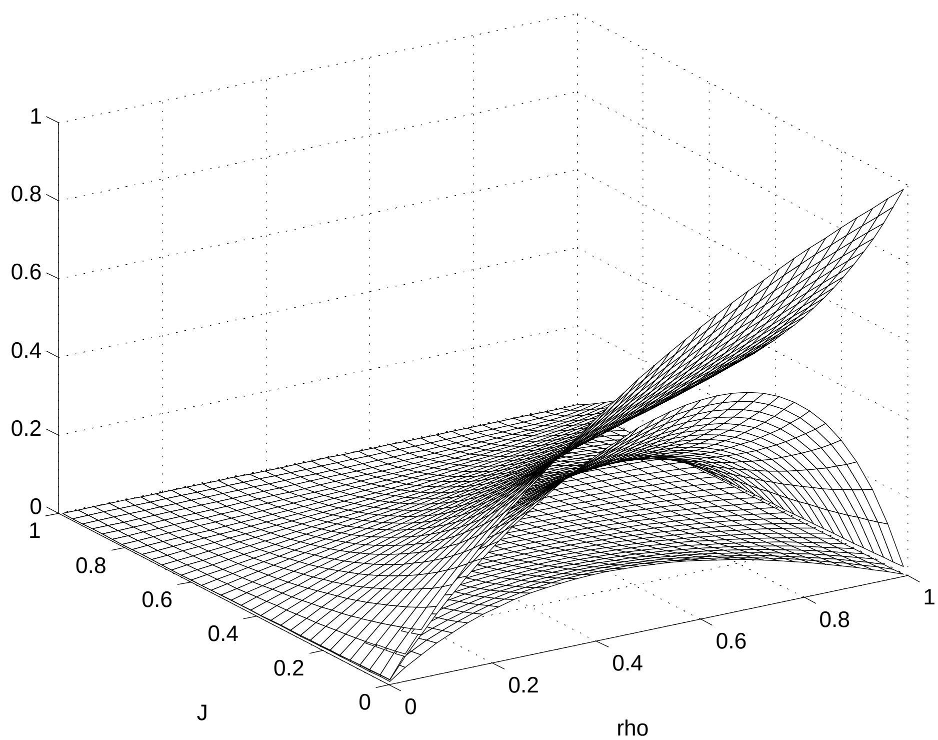

Figure 1: Ratio of variance weighting factors, $f(J, \rho, n)$, for $n = 2, 10, \infty$

Several properties of the variance estimator are discernible from Figure [1]. Most obviously, because $f \leq 1$ for $\rho \in [0, 1]$ and all n, taking ρ to be 0 has a modest effect on the value of the variance, an effect most pronounced for moderate values of ρ (i.e., close to neither 0 nor 1), small values of J (i.e., near 0) and large values of n. Given that f was defined predicated on the assumption of a nearly-degenerate distribution for P, Figure [1] in fact depicts the most extreme variance inflation possible, so that unimodal P distributions would produce substantially muted effects.

Based on Figure [1], the relative effects of the two types of first-order carry-over represented by ρ and J can be compared. Perhaps most obvious is the relative order of the inflationary effects allowed by ρ and J. Where carry-over in the θ series ('baseline' choice probabilities), measured by ρ, has in the most extreme case an effect that causes variance inflation by at most a factor of 2 (for n large, ρ extreme, J near 0 and P near degeneracy), the asymptotic relation to sample size remains on the order of n^{-1}; such effects can be reasonably termed moderate, and ignoring them entirely, in application to actual data, would be unlikely to cause a great deal of overconfidence in parametric convergence, as measured by the 'tightness' of the resulting distribution. By contrast, 'exogenous' (to the utility specification) first-order effects, represented by J, affect the asymptotic rate of convergence itself, with 'large' values of J inflating the rate to an order of n, so that the variance is essentially independent of n. Analogous to the treatment for ρ above, we can define $g(J, \rho, n)$ as the ratio of the variance expressions of the form (19), with the unrestricted expression normed by that when $J = 0$:

$$g(J, \rho, n) = \frac{\lim_{t \to \infty} Var\left[\overline{Y}\right]}{\lim_{t \to \infty;\ J \to 0} Var\left[\overline{Y}\right]} \tag{22}$$

This function, $g(J, \rho, n)$, is depicted in Figure [2] for $n = 2, 5, 10$.

Figure [2] depicts the salient features of the effects of state dependence. For large values of J, the variance inflation factor is quite large, approaching n for extreme values of ρ; in fact, it is not difficult to show that, for $J \approx 1$, the limiting contour apparent in Figure [2] is given by

$$\lim_{J \longrightarrow 1} g(J, \rho, n) = n\,(1+\rho)\left[(1+3\,\rho) - \frac{2\rho\,(1-\rho^n)}{n\,(1-\rho)}\right]^{-1}, \tag{23}$$

which approaches n for ρ near 0 or 1, and reaches a unique minimum between these values.[3]

Thus, the effects of state dependence can be said to be not only far stronger, but of a different nature, than those of habit persistence. Whereas the effects of habit persistence are at best modest, bounded by a factor of 2, those of state dependence increase (nearly linearly, for large J) with the sample size, causing the variance to fall off at a lesser asymptotic rate. Further, while the effects of

[3]No closed form solution exists as a function of n, although it is possible to show that $n/2 \leq \min_\rho g(1, \rho, n) \leq n\left(4 - 2\sqrt{2}\right)^{-1}$ for all n.

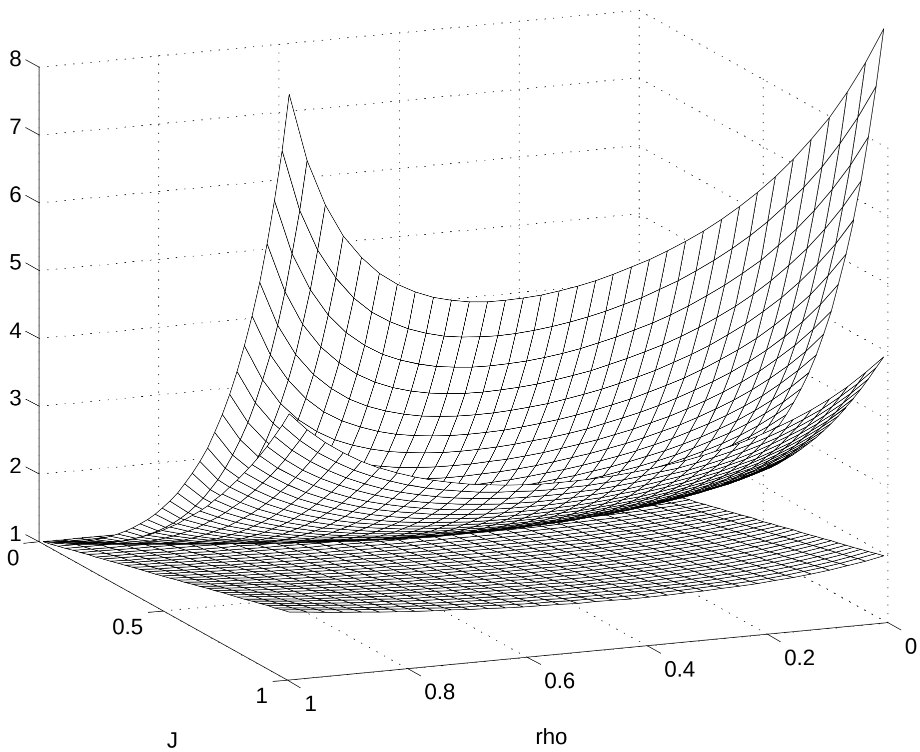

Figure 2: $g(J, \rho, n)$ for $n = 2, 5, 10$

habit persistence are maximized for intermediate values of the parameter ρ (and fall to insignificance near its boundaries) and large values of J, state dependence effects are maximized for large values of the relevant parameter (J) an effect that is partly mitigated for intermediate quantities of ρ. In tandem, while the effects operate synergistically, they do not do so in a separable fashion, with the greatest joint effect dependent on the particular degree of state dependence (generally large) and habit persistence (generally intermediate) in the system.

5 Conclusions

From the earliest efforts in household-level brand choice modeling, it has been clear that some type of feedback mechanism is desirable, if not required, for two reasons: as a brand-intercept heterogeneity correction (e.g., Guadagni and Little 1983; Krishnamurthi and Raj 1991), and to account for preference non-stationarity (e.g., Fader and Lattin, 1993). Such 'constructed' variables are appealing not only for the quality of disaggregate information that they convey, but for their ease of calculation, interpretation and parsimony, an ease which must be traded-off against their possible confounding effects (e.g., intrinsic preference, price sensitivity, carry-over, etc.) when compared with Bayesian methods (e.g., Allenby and Lenk 1994; Lenk and DeSarbo 1996). However, the small-sample statistical properties of such variables have gone largely unexplored, notably their effects on convergence, for example, of baseline household preference measures of the type explored in the present paper.

We have focused on two methods of accommodating carry-over, the first based on habit persistence (temporal autocorrelation) in the choice model, the second in the form of correlations between present choice probabilities and past choices, or state dependence. Their effects, with regard to mean convergence rates, are rather different, not only in relative strength but where in the joint (J, ρ) parameter space they are most pronounced. Compared with those of state dependence, the effects of habit persistence are rather mild, and are at their strongest when there is little or no state dependence present.

The basic methods presented here can be extended from the bivariate case to the multivariate, as discussed previously, by considering the Bernoulli choice as between a focal brand and all other brands, although this follows trivially only in the case of models that conform to the IIA assumption, such as Logit or any model consistent with Luce (1959) type scaling. Further work might entail the study of so-called 'loyalty' variables and their attendant properties; as these are merely linear combinations of past-purchase indicator variables, the implicit $1/n$ weightings of the mean vector would be replaced by an appropriately-normed inner product, so that first and higher moments should in principle follow from the same type of conditioning methods developed here. As this entails, in the Guadagni and Little (1983) treatment, an additional 'smoothing' parameter be estimated, such a derivation has not been undertaken here, though in broad outline the qualitative implications would doubtless be of a similar character with regard to the relative effect strength of state dependence and habit persistence on convergence of the loyalty estimator.

References

[1] 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.

[2] Bawa, Kapil (1990), "Modelling Inertia and Variety-Seeking Tendencies in Brand Choice Behavior," *Marketing Science*, Summer, v9(3), 263-278.

[3] 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.

[4] 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.

[5] Feinberg, Fred M. and Gary J. Russell (1997), "A Consistent Loyalty Measure for Generalized Logit Models," Working Paper, Faculty of Management, University of Toronto.

[6] 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.

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

[8] 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

[9] Hutchinson, J. Wesley (1986), "Discrete Attribute Models of Brand Switching," *Marketing Science*, 5 (Fall), 350-371

[10] Jeuland, Abel P. (1979), "Brand Choice Inertia as One Aspect of the Notion of Brand Loyalty," *Management Science*, 25, 7 (July), 671-682

[11] 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.

[12] Krishnamurthi, Lakshman, and S.P. Raj (1991), "An Empirical Analysis of the Relationship Between Brand Loyalty and Consumer Price Elasticity," *Marketing Science*, 10(2), 172-183.

[13] Lenk, Peter J. and Wayne DeSarbo (1996), "Bayesian Inference of Finite Mixture, Random Effects, Generalized Linear Models," Working Paper, The University of Michigan Business School.

[14] Luce, R. Duncan (1959), Individual Choice Behavior: A Theoretical Analysis, New York, NY: Wiley and Sons, Inc.

[15] Resnick, S. I. and R. Roy (1990), "Multivariate Extremal Processes, Leader Processes, and Dynamic Choice Models," *Advanced Applied Probabilities*, 22, 309-321.

[16] Roy, Rishin, Pradeep K. Chintagunta and Sudeep Haldar (1996), "A Framework for Investigating Habits, "The Hand of the Past," and Heterogeneity in Dynamic Brand Choice," *Marketing Science*, Winter, v15(3), 280-299.