Stochastic Prediction in Multinomial Logit Models

Arthur Hsu • Ronald T. Wilcox
Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
arth@andrew.cmu.edu • wilcox@andrew.cmu.edu

It is standard practice to form predictions from multinomial logit models by ignoring the estimation error associated with the parameter estimates and solving for the predicted endogeneous variable (market share) in terms of the exogenous variables and the point estimates of the parameters. It has long been recognized in the econometrics literature that this type of nonstochastic prediction, which ignores the sampling distribution of the parameter estimates, leads to incorrect inferences about the endogenous variable. We offer a simulation-based approach for approximating the exact stochastic prediction. We show that this approach provides very accurate approximations with minimal computation time and would be easy to implement in industrial applications.
(*Choice Models; Econometric Modeling*)

1. Introduction

It is standard practice to form predictions from multinomial logit models by ignoring the estimation error associated with the parameter estimates and to solve for the endogenous variable as a function of the exogenous variables and the point estimates of the parameters. Given some competitive set of N products, the predicted market share $\hat{S}_k(\hat{u})$ for some product $k \in \{1, \ldots, N\}$ would be computed as

$$\hat{S}_k(\hat{u}) = \frac{e^{\hat{u}_k}}{\sum_{i=1}^{N} e^{\hat{u}_i}}, \tag{1.1}$$

where $\hat{u} = (\hat{u}_1, \ldots, \hat{u}_N)$ is the vector of utility point estimates of the respective products. Because these estimates are obtained from finite samples, the corresponding market share predictions are sensitive to the sampling variation in these utility estimates.

Forming predictions from (1.1), or an analog thereof, is what Klein (1971) termed "deterministic prediction." He instead advocated "stochastic prediction," "where we fully recognize the influence of random error on our judgement." This insight led to a series of papers, mostly by macroeconomists concerned with economic prediction (e.g., Fair 1980), where the sampling distributions of the parameter estimates were incorporated in the prediction model. Dealing more formally with the differences between deterministic and stochastic prediction, Mariano and Brown (1983) showed in a general nonlinear setting that using deterministic predictors yield asymptotically biased and inefficient forecasts even when parameter estimates are consistent, and that the asymptotic mean squared prediction error for stochastic predictors is less than that of deterministic predictors. Similarly, Blattberg and George (1992) showed that using parameter point estimates from one model as inputs into a subsequent decision model (e.g., an estimated demand model as an input into a profit maximization problem) may also lead to suboptimal decisions.

The intuition for these results arises from the well-known fact that $E[g(x)] \neq g(E[x])$ except in very special cases (e.g., g is linear). Therefore, substituting the expected values of the parameters, which are themselves random variables, into a nonlinear function to form a prediction will *not* yield the expected

0025-1909/00/4608/1137$05.00
1526-5501 electronic ISSN

value of the function. In our particular application, multinomial logit (MNL) models, it is generally the case that the market share of small share brands will be underestimated while those of dominant brands will be overestimated by deterministic prediction.

The distinction between deterministic and stochastic prediction arises only in a neoclassical, or frequentist, estimation context. The Bayesian practice of reporting predictive distributions instead of point estimates circumvents the deterministic prediction problem.[1] Yet, maximum likelihood methods, for better or worse, remain the standard in software packages which estimate multinomial logit models. Fn1

The problems associated with deterministic prediction are all the more troubling for marketing researchers. Models estimated with scanner data are often used to make predictions about market share under various marketing mix scenarios. Logit models used in conjoint market share simulators are used to predict a new product's market share given different potential new product designs. Further, there is an increasing trend in marketing towards estimating choice models at lower levels of aggregation (Rossi and Allenby 1993) as well as estimating models which are more structural in their representation of consumer behavior (Erdem and Keane 1996, Gönül and Srinivasan 1996). These trends have led to models with a greater number of parameters and hence greater dispersion in the sampling distributions of those parameters.

The most straightforward way to develop a stochastic predictor for a multinomial logit model is to directly evaluate the expression

$$\hat{S}_k = E_B[\hat{S}_k(\beta_k, x_k)]$$

$$= \int \cdots \int \frac{e^{\beta_k x_k}}{\sum_{i=1}^{N} e^{\beta_i x_i}} f_B(\beta) d\beta, \qquad (1.2)$$

where $f_B(\beta)$ is the joint *p.d.f.* of the parameter estimates for the products in the competitive set and x_k is the vector of attribute values (e.g., the value of the marketing mix variables) for product k.[2] In most scanner data and conjoint applications of the logit model, $f_B(\beta)$ is specified as a multivariate normal density since sample sizes are assumed large enough to invoke the asymptotic normality of the estimates. Even though the multinomial logit function and multivariate normal *p.d.f.* have closed form expressions, (1.2) cannot be evaluated analytically and must be approximated by using a numerical integration technique such as Gaussian quadrature. Similarly, bootstrapping approaches, which generate the empirical sampling distribution of the estimates without resorting to asymptotic approximations, are a reasonable approach to stochastic prediction when the researcher has reason to believe that asymptotic theory provides a poor guide (i.e., small sample size). Yet, both numerical integration and bootstrapping require considerable computation time to evaluate and hence are difficult to implement in popular software packages where solution speed is an important issue.[3] Faced with an expression that is impossible to evaluate analytically and difficult to evaluate numerically, scanner data research and conjoint research have chosen to ignore this problem and proceed with deterministic prediction. Fn2 Fn3

This research provides a practical solution to this problem. We develop a simple and computationally efficient approach to implementing a simulation-based solution to the theoretically correct (stochastic) market share prediction in a multinomial logit framework. We show that this approach would be very easy to implement in industrial applications while closely approximating other, more computationally intensive stochastic prediction methods. First, however, to fix the intuition for the problem at hand, we set out a simple example.

Consider a marketplace where there are two com-

[1] For some examples of marketing applications utilizing Bayesian estimation, see Arora et al. (1998) and Rossi et al. (1996).

[2] In many scanner data applications the response parameters β are assumed to be constant across different choice alternatives.

[3] We note that some recent marketing material distributed by Sawtooth Software, Inc., developer of several popular conjoint analysis packages, indicated that it was not planning on widely distributing an updated, and in many respects better, version of one of its software modules because for many applications the solution time would be 20–30 seconds slower.

peting products ($p = 1, 2$), each defined by a set of attributes. Through observing consumers' choices a market researcher has been able to estimate the parameters of the model and compute the logit utilities Fn4 of these products.[4] Define $\hat{u}_p$ as the utility of product p and Σ_u as the 2×2 variance-covariance matrix of the utility estimates. Let

$$\hat{u}_1 = 3,\ \hat{u}_2 = 4, \quad \text{and} \quad \Sigma_u = \begin{bmatrix} 0.16 & -0.1 \\ -0.1 & 0.25 \end{bmatrix}.$$

Given these parameter estimates, the conventional way to predict the market share of Product 1 would be to follow (1.1) and evaluate

$$\hat{S}_1 = \frac{e^3}{e^3 + e^4}, \tag{1.3}$$

which yields $\hat{S}_1 = 0.2689$ which we will refer to as the *deterministic prediction*. Stochastic prediction for this estimated model requires evaluating

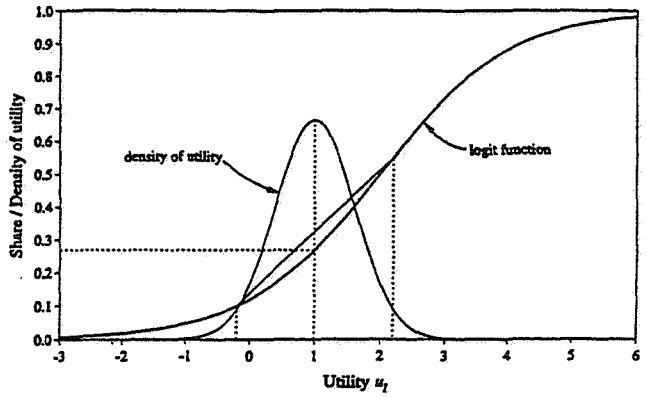
$$\hat{S}_1 = \int_{u_1} \int_{u_2} \frac{e^{u_1}}{e^{u_1} + e^{u_2}} f(u_1, u_2) du_1 du_2, \tag{1.4}$$

where $f(u_1, u_2)$ is the bivariate normal density of the utility estimates. Evaluating this expression by numerical integration yields $\hat{S}_1 = 0.2920$ which shows that the deterministic prediction has a difference of about 0.0231 or 2.31% in total market share. This is a significant discrepancy relative to the 29.2% market share prediction.

The intuition for why this discrepancy occurs is F1 captured in Figure 1. The logit function depicted in this figure was generated by setting $u_2 = 4$ and calculating the logit market share prediction for Product 1 over various values of u_1. Also depicted is the normal density $\mathcal{N}(3, 0.16)$ corresponding to the sampling distribution of u_1 in the example. Deterministic prediction involves evaluating the logit function at the value of u_1 corresponding to the expected value of its sampling distribution (in this case, $u_1 = 3$). Stochastic prediction determines the expected value of the logit

[4] In §3.2, we will lay out how to move from the means and covariance matrix of the parameter estimates β to the means and covariance matrix of the estimated utilities u.

Figure 1 Intuition for the Two-Product Example



function over the entire sampling distribution of u_1. Because the normal distribution is symmetric about u_1 and the logit function is locally convex when $u_1 < 4$, Jensen's Inequality suggests that the deterministic prediction will be less than the stochastic prediction. Likewise, if the point estimate of u_1 is such that $u_1 > 4$, deterministic prediction would produce a market share greater than that of stochastic prediction.

2. A Monte Carlo Approach to Stochastic Prediction

2.1. Overview

Monte Carlo integration is a simulation-based approach to evaluating integrals. The standard Monte Carlo approach involves taking random draws from the joint sampling distribution of the parameters, evaluating the objective function (in this case, the market share of each of the competing products) at each of these draws, and finally taking the arithmetic mean of these evaluations. This approach has been refined through a number of variance reduction techniques (VRT) aimed at reducing the number of draws necessary to achieve a given level of precision. One of these techniques, which is based on the use of antithetic variates, is particularly well suited to the application at hand and we will implement it as a way of increasing the efficiency of our Monte Carlo approach. Before we lay out this approach we show how to exploit some well-known properties of the parameter

estimates to reduce the dimensionality of the simulation problem.

2.2. Transforming the Sampling Distribution of the Parameters to the Sampling Distribution of the Utilities

The computation time associated with numerically evaluating (1.2) via Monte Carlo integration is approximately linear in the dimensionality of the parameter space. We can reduce the dimensionality of this computation from R, the number of parameters, to N, the number of competitive products, by exploiting the asymptotic normality of the parameter estimates as well as some standard properties of variances and covariances. In most applications, the number of parameters will exceed the number of products in the competitive set and so this transformation will lead to a more efficient formulation. Define $\hat{u}_k = \sum_{r=1}^{R} \hat{\beta}_{rk} x_{rk}$ as the point estimate of the utility for product k.[5] Given $\beta \sim {}^{a}\mathcal{N}(\hat{\beta}, \Sigma_\beta)$ this directly implies that the utility estimates are distributed according to $u \sim {}^{a}\mathcal{N}(\hat{u}, \Sigma_u)$ where the diagonal elements of the covariance matrix Σ_u are given by

$$\mathrm{Var}(u_k) = \sum_{r=1}^{R} x_{rk}^2 \,\mathrm{Var}(\beta_{rk}) + 2 \sum_{r=1}^{R} \sum_{r'<r,}^{R} x_{rk} x_{r'k} \,\mathrm{Cov}(\beta_{rk}\beta_{r'k}), \qquad (2.1)$$

and the off-diagonal terms, the covariances between the estimated utilities of any two products $k, k', k \neq k'$ are given by

$$\mathrm{Cov}(u_k u_{k'}) = \sum_{r=1}^{R} \sum_{r'=1}^{R} x_{rk} x_{r'k'} \,\mathrm{Cov}(\beta_{rk}\beta_{r'k'}). \qquad (2.2)$$

Depending on the particular application, reducing the dimensionality of problem in this fashion can significantly reduce the computation time for generating accurate stochastic predictions. In highly parameterized models, this dimension reduction technique may be critical for practical computation of the stochastic predictions. From here forward we will refer to the distribution of the utilities directly instead of the parameters themselves.

2.3. Monte Carlo Simulation with Antithetic Variates (MCAV)

Formally, we take D draws, denoted by $u^{(d)}$, $d = 1, \ldots, D$ from a multivariate normal distribution. $u^{(d)} \sim \mathcal{N}(\hat{u}, \Sigma_u)$, where $\hat{u}$ is the vector of utility point estimates and Σ_u is the covariance matrix of these estimates as derived in the previous section. We then construct D antithetic variates $u^{(\tilde{d})}$ from $u^{(\tilde{d})} = 2\hat{u} - u^{(d)}$. Notice that in MCAV each draw $u^{(d)}$ is used twice, once directly, and again in the form of its antithetic variate. Define

$$S_k^{(d)} = \frac{e^{u_k^{(d)}}}{\sum_{i=1}^{N} e^{u_i^{(d)}}} \quad \text{and} \quad S_k^{(\tilde{d})} = \frac{e^{u_k^{(\tilde{d})}}}{\sum_{i=1}^{N} e^{u_i^{(\tilde{d})}}} \quad \text{for } d = 1, \ldots, D. \qquad (2.3)$$

Define $\hat{I}_k$ as the expected value of market share for product k. Computing this expectation by MCAV proceeds by

$$\hat{I}_k = \frac{1}{D} \sum_{d=1}^{D} \tfrac{1}{2}(S_k^{(d)} + S_k^{(\tilde{d})}). \qquad (2.4)$$

The theoretical advantage of MCAV over standard Monte Carlo sampling is evident from examining the variance of $\hat{I}_k$. Given that the D draws are i.i.d., in standard Monte Carlo sampling $\mathrm{Var}(\hat{I}_k) = \mathrm{Var}(S_k^{(1)})/D$, since $\mathrm{Var}(S_k^{(1)})$ is the variance of any draw. Because of the way that the antithetic variates are formed, $\mathrm{Var}(S_k^{(d)}) = \mathrm{Var}(S_k^{(\tilde{d})})$ for any d and because the draws are i.i.d., $\mathrm{Cov}(S_k^{(d)} S_k^{(f)}) = \mathrm{Cov}(S_k^{(d)} S_k^{(\tilde{f})}) = \mathrm{Cov}(S_k^{(\tilde{d})} S_k^{(\tilde{f})}) = \mathrm{Cov}(S_k^{(\tilde{d})} S_k^{(f)}) = 0$ for any $d \neq f$. Thus, it is straightforward to derive $\mathrm{Var}(\hat{I}_k)$ for MCAV as

$$\mathrm{Var}(\hat{I}_k) = \frac{\mathrm{Var}(S_k^{(1)}) + \mathrm{Cov}(S_k^{(1)} S_k^{(\tilde{1})})}{2D}. \qquad (2.5)$$

Even if the covariance between a draw and its antithetic variate is zero (i.e., $\mathrm{Cov}(S_k^{(1)} S_k^{(\tilde{1})}) = 0$) $\mathrm{Var}(\hat{I}_k)$ is reduced by a factor of two. This is not surprising since

[5] Most high involvement choice models are derived from Luce's Axiom (1959). In these derivations, the "utility" in Luce's model is linked to the parameters and variables in exactly the same fashion as shown here.

the logit function S_k is evaluated twice for each draw. The real benefit of MCAV stems from the negative correlation that is usually induced between each draw and its associated antithetic variate. This negative correlation gives rise to negative covariances between associated market share evaluations (i.e., $\mathrm{Cov}(S_k^{(1)}S_k^{(\tilde{1})}) \leq 0$) and hence, directly from (2.5), reduces $\mathrm{Var}(\hat{I}_k)$.[6] Fn6

2.4. Model Validation

In §§2.1–2.3 we detailed an implementable method for computing stochastic predictions in a multinomial logit model. We now wish to validate this approach in terms of its predictive accuracy in industrial settings. Rather than present results from carefully chosen datasets that strongly support our approach, we construct a set of experiments designed to explore the predictive accuracy of this stochastic prediction over a wide range of possible parameter values.

The parameter values chosen for these experiments are intended to reflect those values typical of industrial applications. We have derived parameter values from publicly available datasets as well as other private datasets and we believe that the parameter values in the experiment detailed below are representative of those found in practice.

To compare the approach that we ultimately advocate with other competing approaches, we simulated the underlying product attribute data (x), not the utilities directly, so that the response parameters (β) could be estimated for different simulated data sets. We simulated five attributes and one brand-specific constant for each brand. Since the scale of attribute variables and estimated response parameters varies widely in industrial data sets we report the distribution of the resulting estimated utilities and utility covariance matrices. That is, we simulated the underlying data such that, after estimating the response parameters and performing the dimension reducing transformations in §2.2, the resulting estimated utilities and utility covariance matrices follow the distributions described below.

Specifically, we designed an experiment consisting of a three-product market. Thus, for each scenario we estimated 17 parameters.[7] We generated 80 three-product scenarios following the parameter distributions described below. After estimating the parameters, each scenario provided a vector of aggregate product utility estimates and an estimated matrix of utility covariances. Of these scenarios, 40 contained 100 sample observations while the remaining 40 contained 1,000 sample observations. Because the recommendations we detail later rely on the asymptotic properties of the estimates, choosing a large sample size of 1,000 is conservative since almost all industrial data sets, including relatively small conjoint data sets, are at least this large. Our experimental design is as follows. Fn7

- The utility estimates for the three products are u_i, $i = 1, 2, 3$. In our experience, the range of differences between the utility estimates of products within a competitive set does not exceed six. Since the market share prediction is driven not by the value of the utility estimates themselves but rather by the difference in utility values $u_i - u_j$, for products i, j within a competitive set, we constructed the computational study such that the utility estimates range between zero and six as follows: $u_i \underset{i.i.d.}{\sim} U[0, 6]$, where U denotes the uniform distribution.[8] Fn8

- The variances of the utility estimates for product i are σ_{ii}. In our experience, most variances of the utility estimates are less than 0.5 while a few range as high as 1.5. Consistent with this experience we chose a distribution for the variances such that 80% of the density lies uniformly between 0 and 0.5 with the remaining 20% lying uniformly in the interval 0.5 to 1.5. It follows that the density function of the variance of the estimated utilities is

$$f(\sigma_{ii}) = \begin{cases} 1.6, & 0 \leq \sigma_{ii} < 0.5, \\ 0.2, & 0.5 \leq \sigma_{ii} < 1.5, \\ 0, & \text{otherwise.} \end{cases}$$

- The covariances between the utility estimates for products i and j are σ_{ij}, $i \neq j$. Our experience is not much

[6] For a detailed description of antithetic variates, see Law and Kelton (1991, pp. 628–634).

[7] The number of parameters is equal to the number of products multiplied by the number of attributes per product less one brand-specific constant. In our application $(3 * 6) - 1 = 17$.

[8] The range selected for the utility values is sufficient to allow market shares to range between 0.2% to 99.5%.

Table 1 Results of Monte Carlo Experiments

Predictions	Mean Rel. Error–SS	Mean 95% C. I. Half-width–SS	Mean Rel. Error–LS	Mean 95% C. I. Half-width–LS	CPU seconds
Deterministic	9.60%	*	6.63%	*	*
MC(1,000)†	0.74%	$3.17 * 10^{-3}$	0.10%	$5.96 * 10^{-4}$	0.16
MC(100,000)	0.58%	$3.18 * 10^{-4}$	0.05%	$5.95 * 10^{-5}$	15.74
MCAV(1,000)	0.58%	$2.25 * 10^{-3}$	0.03%	$4.20 * 10^{-4}$	0.02
MCAV(100,000)	0.52%	$2.25 * 10^{-4}$	0.02%	$4.21 * 10^{-5}$	3.60
Numerical Integration	0.83%	*	0.70%	*	340.34
Bootstrapped	*	*	*	*	1834.84

† MC(D) refers to prediction using Monte Carlo method with D draws while MCAV(D) describes the Monte Carlo method with antithetic variates.

of a guide in setting these parameter values. We have observed both positive and negative covariances in scanner datasets. We have often observed covariances close to zero in conjoint datasets where orthogonal (D-optimal) experimental designs tend to produce uncorrelated utility estimates. Since these covariances may vary widely across datasets we allow the covariance parameters in our computational experiment to vary widely as well. Specifically, σ_{ij}, $i \neq j$ is distributed as follows: $\sigma_{ij} \underset{i.i.d.}{\sim} U[-\sqrt{\sigma_{ii}\sigma_{jj}}, \sqrt{\sigma_{ii}\sigma_{jj}}]$ with $\sigma_{ji} = \sigma_{ij}$ and subject to the restriction that the resulting covariance matrix Σ_u is positive semidefinite. This range spans the space of possible covariances since any σ_{ij} outside this range would violate the positive semidefinite requirement of a covariance matrix.

With three products for each of the 80 scenarios, this experiment resulted in 240 market share observations for which we can examine market share predictions generated by: (1) the deterministic method, (2) Monte Carlo integration, (3) MCAV with the dimension-reducing transformation, (4) numerical integration by Gaussian quadrature, and (5) bootstrapped sampling distributions. The results T1 are summarized in Table 1 where "SS" indicates results for the small sample scenarios, and "LS," the large sample scenarios.

We define relative error as the ratio of the difference between the predicted market share and the stochastic prediction obtained via bootstrapping to the bootstrapped stochastic prediction. We report the mean relative error, with the mean taken over the 120 market share observations for each sample size. In each of the Monte Carlo experiments, it is possible to obtain not only the point estimate of the market share, but also a confidence interval on the estimate. Given the sample variance S^2 over the D draws, the half-width of the confidence interval centered around the point estimate at an α level of confidence is $t_{D-1,1-\alpha/2}\sqrt{S^2/D}$. Finally, we report all mean computation times in CPU seconds for a three-product scenario measured on a Sun SPARCstation 20/61.[9] Fn9

It is clear from Table 1 that as we increase the number of draws, the MCAV method provides substantial improvements in predictive accuracy at a very modest computational cost. For example, in our large sample experiment with 100,000 MCAV draws we decrease the mean relative error from 6.63% to 0.02% compared with deterministic prediction. This gain comes at the cost of only 3.60 seconds of computation time. While we feel that 100,000 draws imposes a very minimal computational burden on the user, note that the users themselves can make the trade-off between computation time and accuracy. For the large sample size experiments, the MCAV closely approximates the bootstrapping solution but with a minimal amount of computation time. This directly implies that the normal approximation required by the MCAV procedure

[9] The difference in computation time between the large and small sample experiments was negligible. Therefore, we report only CPU times for the large sample experiments.

Fn10 is not restrictive.[10] As expected, in smaller samples the accuracy of those prediction methods which rely on asymptotic approximations deteriorates. Further, Table 1 indicates that MCAV with the dimension reducing transformation provides benefits over the standard Monte Carlo approach. For this set of experiments, MCAV was on average 8.8 times faster at producing predictions of equal accuracy to standard Monte Carlo.

Finally, the value of stochastic prediction hinges on the variances and covariances of the product utility estimates. Typically, the difference between deterministic and stochastic prediction will increase in the variances of the utility estimates. The variance of the utility estimates themselves increase in the number of parameters present in the model as well as the variances of the individual parameter estimates (2.1). Further, positive covariances among the product utilities tends to dampen the difference in the market share predictions between the two prediction methods while negative covariances tend to exacerbate it. These negative covariances often arise in models where the analyst allows brand-specific coefficients for marketing mix variables (i.e., the price coefficient is allowed to vary across brands) instead of constraining them to be the same across all brands.

2.5. Step-by-Step Guide to Stochastic Prediction for Multinomial Logit Models

Given the above discussion, we believe it is reasonable to recommend the MCAV approach with the dimension reducing transformation for developing stochastic predictions with MNL models in most marketing applications. Only when sample sizes are quite small does moving to the empirical sampling distributions by bootstrapping offer significant benefit. We now lay out a step-by-step procedure for developing stochastic predictions by MCAV.

Step 1. Estimate the parameters of the MNL model along with the covariance matrix of these parameters using maximum likelihood.

Step 2. Compute the estimated product utilities $\hat{u}_k = \sum_{r=1}^{R} \hat{\beta}_{rk} x_{rk}$ where k indexes the products and r indexes the variables included in that product's utility function.

Step 3. Compute the covariance matrix of these utilities Σ_u by substituting the parameter estimates and covariances into Equations (2.1) and (2.2).

Step 4. Take D draws (we recommend at least 10,000) from a multivariate normal distribution, $u^{(d)} \sim \mathcal{N}(\hat{u}, \Sigma_u)$, and save each of these draws.

Step 5. For each of these k-dimensional utility draws compute a k-dimensional antithetic variate by $u^{(\tilde{d})} = 2\hat{u} - u^{(d)}$.

Step 6. For each utility draw and each antithetic variate draw compute a market share evaluation for each brand by substituting the draws into the $S_k^{(d)}$ and $S_k^{(\tilde{d})}$ expressions in Equations (2.3). You will now have $2D$ market share evaluations, where D is the number of original utility draws.

Step 7. Compute the stochastic prediction for the market share of each brand by substituting the market share evaluations computed in Step 6 into Equation (2.4).

3. Summary

The problems associated with deterministic prediction have been largely ignored by many marketing researchers and practitioners. Deterministic prediction ignores the sampling distributions of the parameter estimates and can lead to asymptotically biased and inefficient forecasts.

This paper proposed a practical solution to this problem. We developed a simple and computationally efficient approach to implementing a simulation-based solution to the theoretically correct (stochastic) market share prediction in a multinomial logit framework. This method allows us to circumvent the deficiencies of deterministic prediction at a reasonable computational cost. An experimental study suggested that, within the range of parameter values found in many applications, the MCAV approach rapidly offers substantial increases in predictive accuracy over deterministic prediction. As more disaggregate and structural choice models emerge, we expect our approach to yield even greater benefits.[11]

[10] We performed Kolmogorov-Smirnov tests for normality of the resulting sampling distributions. In our large sample experiment, all bootstrapped sampling distributions were accepted as normal by this test.

[11] We thank Linda Argote, Peter Boatwright, Dennis Epple, Ron Goettler, Rich Johnson, three anonymous reviewers, and the associate editor for comments that materially contributed to this research.

References

Arora, N., G. M. Allenby, J. L. Ginter. 1998. A hierarchical Bayes model of primary and secondary demand. *Marketing Sci.* 17 29–44.

Blattberg, R. C., E. I. George. 1992. Estimation under profit-driven loss functions. *J. Bus. Econom. Statist.* 10 437–444.

Erdem, T., M. P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic choice processes in turbulent consumer goods markets. *Marketing Sci.* 15 1–20.

Fair, R. C. 1980. Estimating the expected predictive accuracy of econometric models. *Internat. Econom. Rev.* 21 355–378.

Gönül, F., K. Srinivasan. 1996. Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Sci.* 15 262–279.

Klein, L. R. 1971. *An Essay on the Theory of Economic Prediction.* Markham, Chicago, IL.

Law, A., W. D. Kelton. 1991. *Simulation Modeling and Analysis.* McGraw-Hill, New York.

Mariano, R. S., B. W. Brown. 1983. Asymptotic behavior of predictors in a nonlinear simultaneous system. *Internat. Econom. Rev.* 24 523–536.

Rossi, P. E., G. M. Allenby. 1993. A Bayesian approach to estimating household parameters. *J. Marketing Res.* 30 171–182.

——, R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* 15 321–340.

Accepted by Dipak C. Jain; received August 15, 1998. *This paper was with the authors* 5 *months for* 2 *revisions.*