Misspecification Testing of Unobserved Effects in Hierarchical Bayes Choice Models

Rick L. Andrews

Andrew Ainslie

Imran S. Currim

Rick L. Andrews is the Greer Chair of Business Administration and a professor, Department of Marketing, Louisiana State University (e-mail: randrews@lsu.edu). Andrew Ainslie is an assistant professor, Anderson School of Management, University of California, Los Angeles (e-mail: andrew.ainslie@anderson.ucla.edu). Imran S. Currim is Corporate Partners Research Scholar and a professor, Graduate School of Management, University of California, Irvine (e-mail: iscurrim@uci.edu). The authors contributed equally to the article.

Misspecification Testing of Unobserved Effects in Hierarchical Bayes Choice Models

Abstract

Hierarchical Bayes choice models are seeing widespread adoption in marketing research, partly because of their ability to generate individual-level parameter estimates with limited data. However, the power of such models may tempt researchers to trust that the models continue to produce reasonable estimates when in fact either model misspecification or insufficient data limit the models' ability to recover individual-level parameters successfully. This study examines the impact of misspecified unobserved effects such as parameter heterogeneity, consideration set effects, and state dependence on parameter recovery, fit, and prediction for logit choice models formulated and estimated from a Bayesian perspective. There is reason to believe that the potential for spurious results is high when these unobserved effects are not specified and possibly even when they are, since such unobserved effects may produce choice persistence whose origin is difficult to determine. An extensive simulation experiment is conducted wherein ten logit model specifications are applied to 405 choice datasets varying in terms of the patterns of parameter heterogeneity, consideration set effects, and state dependence present, among other factors. The results show that fit, prediction, and parameter recovery do not go hand-in-hand and that there is a high probability of spurious effects. Consequently, model selection should be based on managerial objectives for model building, such as marketing mix planning, segmentation, or forecasting. The flexibility inherent in a more complete specification of consideration sets, purchase event feedback, and mixtures of normals parameter heterogeneity is found to result in more accurate parameter estimates, making this model more suitable for marketing mix planning or segmentation. A simpler model specified with state dependence and normally distributed parameter heterogeneity is found to provide the most accurate forecasts. Though the potential benefits of using HB models are clear, researchers should exercise caution when using such models.

Misspecification Testing of Unobserved Effects in Hierarchical Bayes Choice Models

Introduction

A recent study by Abramson, Andrews, Currim, and Jones (2000, hereafter AACJ) presents an extensive simulation experiment that examines the extent of parameter bias in logit models resulting from the misspecification of four unobserved effects—consideration set effects, heterogeneity in preferences and responses to the marketing mix, state dependence, and serial correlation. All four effects are believed to produce persistence in observed choices of frequently-purchased packaged goods and hence may be difficult to disentangle. The results, which show that the potential parameter bias from misspecified unobserved effects in choice models can be quite large, sometimes confirm and sometimes contradict conventional wisdom about the consequences of misspecification in choice models. Importantly, the study shows that it is not possible to infer the quality of parameter estimates, which is unobservable in studies using actual marketplace data, from observable fit and prediction statistics.

An issue for future research identified by AACJ is whether their results also hold for models cast with individual-level parameters, in a Bayesian framework. Hierarchical Bayes (HB) models are seeing widespread adoption in marketing research today as advances in computational algorithms permit increasingly complex model specifications and rich heterogeneity structures. The AACJ study restricted the investigation of misspecified unobserved effects to homogeneous logit models and finite mixture models assuming discrete distributions of heterogeneity for the coefficients, all of which were estimated using classical maximum likelihood techniques. The focus of the present study is to examine the impact of misspecified unobserved effects on parameter estimates, fit, and prediction for logit choice models formulated and estimated from a Bayesian perspective. It is typical in Bayesian estimation to allow for individual-level parameters, since a combination of shrinkage and hierarchical regression structures allows for inferences that are difficult using classical estimation methods. However, the danger is that a researcher may rely overly on these advantages in situations where the data are too sparse at the individual level for inferences to be made correctly. We investigate such situations, in conjunction with model misspecification.

One could perhaps argue that the potential for incorrectly identifying unobserved effects is larger with HB models than with the classical models estimated in the AACJ study, because of the incautious researcher's dependence on these advantages of Bayesian methods. For an example of a careful analysis of such a situation, consider the study by Chiang, Chib, and Narasimhan (1999), who develop an HB model that allows heterogeneity in preferences, marketing mix responses, and consideration sets. Using scanner panelists' purchases of ketchup, they estimate logit models (i) with no parameter or consideration set heterogeneity, (ii) with parameter heterogeneity but not consideration set heterogeneity, (iii) with consideration set heterogeneity but not parameter heterogeneity, and (iv) with both parameter and consideration set heterogeneity. Though their study uses real data in which the true effects are unknown, their findings suggest considerable opportunity for spurious effects in HB-estimated logit models.

The potential for spurious effects are of substantial concern because managers often use parameter estimates for market segmentation or assessment, to judge the extent to which customers vary in their loyalty to brands or sensitivity to price and promotion, as a precursor to making price and promotion decisions. While econometric theory informs us about the causes of bias, the extent of bias due to misspecification of unobserved effects in typical scanner data based applications is unclear.

The current study describes a large-scale simulation experiment that generates scanner datasets varying on five data characteristics: (i) the number of purchases per household, (ii) the number of households, (iii) consideration set effects, (iv) parameter heterogeneity, and (v) state dependence. A simulation is useful for identifying spurious effects because the true effects are known with simulated data (e.g., Vriens, Wedel and Wilms 1996), unlike real data. Compared to the design used by AACJ, the present simulation does not examine serial correlation, which was the least problematic unobserved effect in that study, but varies instead the number of purchases per household and the number of households, two variables that affect the quality of estimates in HB-estimated models (see Andrews, Ainslie, and Currim 2002). Ten models are estimated for each dataset, varying in complexity from a homogeneous logit model explaining no unobserved effects to an HB-estimated model explaining parameter heterogeneity formed by mixtures of normal distributions, state dependence, and heterogeneous consideration set effects. Measures of model performance include the extent of parameter estimation error, fit, and prediction accuracy.

The plan of this study is as follows. First, we review the evidence on the incorrect identification of unobserved effects in logit choice models. We then present the design of the simulation experiment, including the data conditions, models estimated, and measures of model performance, followed by the results, conclusions, and implications for marketing modelers and managers.

Background

The AACJ study provides some intuition as to why unobserved effects might be difficult to disentangle with logit models applied to scanner panel data: they all can result in persistence in observed choices. For example, a consumer may make repeat purchases of a brand because previous choice outcomes affect current choices (state dependence), because the consumer has strong preferences for the brand (parameter heterogeneity), or because the consumer did not consider any other brands (consideration set). When models do not correctly identify the cause or causes of the persistence in choices, spurious state dependence, parameter heterogeneity, or consideration set effects may occur, and parameter estimates, fit, and prediction accuracy can be affected.

The results of the AACJ study show that underspecified state dependence and choice set effects may result in substantial parameter bias, while underspecified parameter heterogeneity and serial correlation typically do not. They also find that underspecified state dependence results in spurious parameter heterogeneity but not vice-versa; that underspecified consideration sets result in exaggerated state dependence and vice-versa; that underspecified choice set effects result in exaggerated parameter heterogeneity but not vice-versa; that correctly-specified models recover parameters reasonably well, except when there are multiple unobserved effects in the data, in which case even the most suitable models have difficulty recovering parameters; and that it is difficult to infer parameter bias on the basis of predictive accuracy. As mentioned, the AACJ study did not examine any models that were formulated and estimated from a

Bayesian perspective, and because Bayesian estimation generally incorporates richer structures than its classical maximum likelihood counterpart, it is unclear if we can extrapolate the results of the AACJ study to Bayesian models.

The study by Chiang, Chib, and Narasimhan (1999) proposes an HB model that is capable of accounting for the heterogeneity in consideration sets and in the parameters of the brand choice model. Applying the model to scanner panel purchases of ketchup, they find that ignoring consideration set heterogeneity understates the impact of the marketing mix and overstates the impact of preferences and state dependence, even when heterogeneity in parameters is modeled; that when consideration set heterogeneity is modeled, the parameter heterogeneity takes on less importance; and that the estimate of consideration set heterogeneity is robust to the inclusion of parameter heterogeneity. They include a lagged purchase variable in their model specification but do not investigate the effects of removing the lagged purchase variable on consideration set heterogeneity and parameter heterogeneity. Also, since they use actual scanner panel data, the true effects in the data are not known so that the presence and extent of spurious effects cannot be confirmed. In addition, any study using real data can be viewed as a single cell in a multiple cell simulation experiment. Still, the Chiang, Chib, and Narasimhan study provides a strong indication that misspecification of unobserved effects with HB models could have serious consequences.

Other studies have compared the performance of HB models with finite mixture (FM) models. For example, Andrews, Ansari, and Currim (2002) compare the performance of HB and FM conjoint models, varying aspects of parameter heterogeneity in the data (e.g., number of segments, separation between segments, variation within segments, etc.) as well as sample size and error variance. Andrews, Ainslie, and Currim (2002) perform a similar study for logit models, varying similar factors. Both studies conclude that there is little difference in the accuracy of parameter estimates generated by HB and FM models, though the Andrews, Ainslie, and Currim (2002) study does show that HB models with individual level effects do not perform well when the model is poorly identified at the household level. Neither of these studies examines data or models with state dependence and consideration sets, and so they are unable to address the issue of interest in the current study, which is the potential for incorrect identification of unobserved effects in logit models cast in a Bayesian framework. Wedel et al. (1999) compare HB and FM methods at a conceptual level.

Recent developments in computational methods and the expanded availability of detailed marketplace data have fueled the growth in applications of Bayesian methods in marketing (Rossi and Allenby 2003). With Bayesian methods now becoming widespread in the marketing literature, resulting in increasing numbers of models estimated with individual-level parameters, now is the time to assess the potential for misspecification and misleading managerial insights, not after hundreds of applications have accumulated in academia (and perhaps thousands in industry). Consequently, this study could make a major contribution to the art and science of modeling scanner panel choice data with Bayesian logit models by identifying which unobserved effects must be modeled and which may be ignored without risk of creating spurious effects and substantial parameter bias and by determining how model specification decisions may depend on the managerial objectives for model building.

Simulation

In this section, we discuss the simulation design, i.e., data generation, models estimated, and measures of model performance. The levels of factors chosen for data generation are based on past studies and experience with scanner data sets.

Data Generation

Five factors are manipulated in the simulation experiment, as discussed below.

(i) Number of purchases per household: 5, 10, or 15;

(ii) Number of households: 100, 200, or 300

(iii) Consideration set effects: none, all possible sets, and promotion expansion sets (Siddarth, Bucklin, and Morrison 1995);

(iv) Parameter heterogeneity: none, normal with variance .05, normal with variance .25, mixture of normals—fat tails, and mixture of normals—bimodal;

(v) State dependence: none, first-order, and higher-order (Guadagni and Little 1983).

The first two factors manipulate the sample sizes available for analysis and are straightforward to implement. The study by Andrews, Ainslie, and Currim (2002) suggests that the number of purchases per household may be a more critical factor than the number of households. In that study the number of purchases per household is varied at 3, 10, and 15 purchases. In particular, HB-estimated models with individual level parameters do not perform well when there were only three purchases per household, resulting in a model that is unidentified at the household level. The number of households considered in that study are 75, 200, and 400. Vriens, Wedel, and Wilms (1996) consider 100 and 200 households. These considerations influenced the selection of levels for the first two factors.

Factor (iii) manipulates the consideration set effects in the data over three levels: no consideration sets, all possible consideration sets in the data, and consideration sets formed by the promotion expansion strategy described by Siddarth, Bucklin, and Morrison (SBM, 1995). Two of the three levels selected are similar to those considered by AACJ, who also examined the more compensatory strategy suggested by Bronnenberg and Vanhonacker (1996) but found that this strategy did not result in serious bias. With the five brands used in the simulation, consumers may use any of 2^5-1=31 consideration sets in making choices under the all possible consideration sets scenario, though the consumer is assumed to use the same consideration set throughout his/her purchase history. The promotion expansion strategy assumes that the consumer's consideration set consists of previously purchased brands augmented by currently promoted brands, resulting in a dynamic consideration set. Consistent with the AACJ study and previous empirical results, we specify that 25% of households make choices from the full set of available alternatives, while 75% use reduced consideration sets. When households use a reduced consideration set, which set they use is determined randomly.

Factor (iv) creates five different levels of parameter heterogeneity. At the first level, there is no parameter heterogeneity. The second and third levels assume normally distributed parameters across consumers, with variances of .05 and .25, respectively. Andrews, Ainslie, and Currim (2002) consider the same level of variances, while Vriens Wedel and Wilms (1996) consider variances of .05 and .10. The fourth and fifth levels of factor (iv) assume that the heterogeneity is formed by a mixture of two normal distributions with weights .75 and .25. At the fourth level, the means of the two segment parameter vectors are the same, the first has a variance of .10, and the second has a variance of .50, twice the level of variances considered by Andrews, Ainslie, and Currim (2002) and five times larger than the variances considered by Vriens, Wedel, and Wilms (1996), resulting in fat-tailed distributions. At the fifth level, the means of the two segment parameter vectors differ by 1 on average, which is equal to the largest separation considered by Andrews, Ainslie, and Currim (2002), and both segments have variances of .05, resulting in a bimodal distribution of parameters.

Factor (v) generates state dependence at three levels. At the first level, there is no state dependence. First-order state dependence occurs at the second level of factor (v), operationalized as a lagged dependent variable (e.g., Roy, Chintagunta, and Haldar 1996). Higher-order state dependence is operationalized using the Guadagni-Little (GL) loyalty variable, which consists of an exponentially smoothed purchase history with a smoothing constant of .75. Such a smoothing constant is consistent with previous empirical results as well as the AACJ study. Although the GL loyalty variable has limitations (e.g., Fader and Lattin 1993), the AACJ study found that the measure performed very well in capturing higher-order state dependence. Consequently, we include this type of higher-order state dependence in the study.

With the factor levels described above, the simulation generates 3x3x3x5x3=405 datasets overall. The brand choice task is based on a 5-brand market, and the marketing mix variables consist of price and two binary promotional variables (typically store feature advertising and aisle display). Given the marketing mix variables and the parameters defining the choice process (described above), utilities are calculated, and a double exponential error vector is added to the utilities. After computation of choice probabilities using the logit functional form, the consumer is assumed to choose the brand with the highest utility.

Models

Ten models are estimated for each of the 405 datasets, as described below. Following these brief model descriptions, we provide more details on the implementation of the models.

Model 1: Homogeneous logit. This model specification does not explain any of the unobserved effects—only preferences and the effects of the marketing mix variables. Its seven parameters are estimated using maximum likelihood techniques.

Model 2: Homogeneous logit with GL loyalty. With the addition of the GL loyalty variable, this model is equipped to explain both first-order and higher-order state dependence but not parameter (heterogeneity in preferences and the effects of marketing mix) or consideration set heterogeneity. Nine parameters (including the smoothing constant) are estimated using maximum likelihood techniques.

Model 3: HB logit. Cast in a Bayesian framework, this model should be able to explain normally-distributed parameter heterogeneity, but not state dependence or consideration set heterogeneity.

Model 4: HB logit with lagged dependent variable. With the addition of a lagged dependent variable to model 3, this model should be able to explain first-order state dependence (but not higher-order state dependence) and normally-distributed parameter heterogeneity.

Model 5: HB logit with GL. The addition of the GL loyalty variable to model 3 should allow this model to explain both forms of state dependence (first-order and higher-order) and normally-distributed parameter heterogeneity.

Model 6: HB logit with consideration sets. Based on the consideration set formulation described by Chiang, Chib, and Narasimhan (1999), this model should be able to explain normally-distributed parameter heterogeneity and "All possible" consideration set effects; it is not clear whether the formulation will recover consideration sets formed using the promotion expansion strategy.

Model 7: HB logit with mixture of normals. Based on the formulation described by Allenby, Arora, and Ginter (1998), this model should be able to explain richer forms of parameter heterogeneity but not state dependence or consideration set heterogeneity.

Model 8: HB logit with consideration sets and GL. This model should be able to explain "all possible" consideration set effects, all forms of state dependence, and normally-distributed parameter heterogeneity.

Model 9: HB logit with mixture of normals and GL. Adding the GL loyalty variable to model 7 should allow this model to explain any form of parameter heterogeneity and any form of state dependence.

Model 10: HB logit with consideration sets, mixture of normals, and GL. With the most complete model specification, this model should be able to explain any form of parameter heterogeneity, any form of state dependence, and "all possible" consideration set heterogeneity.

Models 10, 9, 8, and 5 have not been estimated in the choice modeling literature; of these 10 models, the AACJ study considered only models 1 and 2. None of the models cast in a Bayesian framework has been subjected to such an extensive simulation-based test to determine which unobserved effects must be modeled and which may be ignored without risk of creating spurious effects.

The modularity of hierarchical models and MCMC estimation methods allows consideration set effects, mixture effects, and state dependence effects to be added and removed without great difficulty. In general, the specification of the priors for the consideration set component is based closely on that used by Chiang, Chib, and Narasimhan (1999), while the specification of priors for the mixture of normals formulation is based closely on that used by Allenby, Arora, and Ginter (1998). When mixtures of normal distributions (Allenby, Arora, and Ginter 1998) were included in models, we assumed that the number of components was two; we did not attempt to determine the correct number of segments empirically. As for the GL loyalty variable, the smoothing constant was estimated by adding another layer (a Metropolis-Hastings algorithm) to the Gibbs sampler. For simplicity, the same smoothing constant was used for both segments when a GL loyalty variable was used in conjunction with a mixture of normal distributions for heterogeneity. Likewise, the same consideration set structure was assumed for both segments—that is, the consumer was assumed to use the same consideration set whether they belonged to segment 1 or segment 2. For the covariance matrix D of the parameters for all of the HB models, we chose an inverse Wishart prior, $D^{-1} \sim W\left[\rho, \left(\frac{1}{\rho}\right) I\right]$, where ρ is the number of explanatory variables plus two, consistent with the specification of Chiang, Chib, and Narasimhan (1999). A proper but uninformative prior was used for the mean of the parameters as well. For each model/dataset combination, we ran the Gibbs sampler for 10,000 iterations, 5,000 for burn-in and 5,000 for inference.

Performance Measures

Like other recent simulation studies (e.g., AACJ; Andrews, Ainslie, and Currim 2002; Vriens, Wilms, and Wedel 1996), we use several measures to assess the performance of the models in terms of parameter recovery error, fit, and predictive accuracy. Following these studies, the Root Mean Squared Error between the actual and estimated household-level parameters, RMSE(β), is computed as a measure of parameter recovery error. The marginal log likelihood is used to assess model fit. In the case of the homogeneous logit models, BIC was used as an approximation to the marginal likelihood. BIC is based on the Schwartz criterion, which in turn is an approximation to the marginal likelihood of a model (Schwartz 1978). For the HB models, the marginal likelihood was computed using the reweighted importance sampling method outlined in Raftery (1996) and Newton and Raftery (1994). In order to make this comparable to the BIC, a statistic widely used in the marketing literature, we compared –2(marginal log likelihood) to the BIC measure for the other models. Hereafter, we refer to both as -2MLL. Finally, we used the log likelihood from the validation sample to assess prediction accuracy.

Results

Table 1 shows the means for the parameter recovery error measure RMSE(β) by model type and factor level. The superscripts on the model-type means at the bottom of the table indicate statistically significant differences, with "1" being the best. Table 2 shows the corresponding results for the fit measure -2MLL, and Table 3 shows the results for the predictive validation measure LOGL-V.

Parameter Recovery and Fit

Examining Table 1, we see that M10 (HB logit with consideration sets, mixture of normals, and GL) has the best RMSE(β) values on average, with approximately 18% smaller parameter recovery error than the homogeneous logit model with GL variable (M2). The RMSE(β) value for M10 is slightly more than half as large as that of a zero-order logit model explaining no unobserved effects (M1). According to the means for -2MLL in Table 2, however, M8 (HB logit with consideration sets and GL) has the best fit in almost all data conditions. Though it may appear that M10 would have to fit at least as well as M8 (indeed this would be true with classical estimation methods), this is not the case with HB estimation methods.

Delving deeper into Table 1, we can see that the consideration set factor in the data is determining the relative performance of the various models. When there are no consideration effects in the data or when consideration is via the strategy described by SBM, the homogeneous logit model with GL loyalty (M2) has the best parameter recovery error, which is quite surprising given that 80% of the datasets have unobserved parameter heterogeneity of some form. However, when all possible consideration sets are used in the data, M10 (HB logit with consideration sets, mixture of normals, and GL) is slightly better than M8 (HB logit with consideration sets and GL), which is slightly better than M6 (HB logit with consideration sets), and the models without consideration sets do not perform as well. With the big advantage M10 has over M2 in the “All possible” consideration sets level (43% smaller parameter recovery error), M10 appears to be best in almost all data conditions in Table 1. Thus, the means for the "Consideration Sets" factor in Table 1 are insightful, but all other factor level means are less so because of the influence of the "All possible" level.

Therefore, Table 4 examines selected RMSE(β) means at a more disaggregated level to gain insights into the effects of underspecifying choice set effects, parameter heterogeneity, and state dependence. For example, rows 1-4 in Table 4 examine the RMSE(β) means for datasets which use (1) all possible consideration sets, (2) all possible consideration sets but no parameter heterogeneity, (3) all possible consideration sets but no state dependence, and (4) all possible consideration sets but no parameter heterogeneity or state dependence. Tables 5 and 6 show the same means for the fit and prediction measures. We discuss these results below.

Consideration Set Effects

From rows 1-4, we can assess the impact of underspecifying consideration sets when all consideration sets are possible in the data. M9 should be able to handle all forms of parameter heterogeneity and all forms of state dependence but not consideration set heterogeneity, while M10 should be able to handle all effects modeled by M9 in addition to "All possible" consideration set heterogeneity. Thus, a comparison of M9 and M10 gives an indication of the consequences of failing to model "All possible" consideration set heterogeneity. As expected, we see that the reduction in RMSE(β) from modeling "All possible" consideration set heterogeneity is extremely large: .26 vs. .80 (or 68% improvement) when there are no other unobserved effects, .45 vs. .89 (or 49% improvement) when there is no state dependence, .28 vs. .71 (or 61% improvement) when there is no parameter heterogeneity, and .47 vs. .83 (or 43% improvement) across all data conditions.

In contrast to the findings for the “All possible” consideration sets condition, we see that none of the models with consideration set specifications (M6, M8, and M10) is more effective in capturing SBM consideration set usage (rows 5-8), leaving M2 (homogeneous logit with GL) as the model with the best parameter recovery. For example, M2 offers improvement in parameter recovery over M6 of 17%, 38%, 11%, and 41% in rows 5, 6, 7, and 8, respectively. We know from the AACJ study that a homogeneous logit model with GL and a model structure consistent with SBM consideration set formation performs quite well and likely would be the best model in the SBM data condition. A heterogeneous version of a logit model with SBM consideration set effects is quite feasible to estimate with HB methods and might be better still.

Given consideration set effects in the data (rows 1-8), we can examine the potential for spurious state dependence and parameter heterogeneity when consideration set effects are underspecified. It appears that inclusion of the GL loyalty variable when there are underspecified “All possible” choice set effects can produce spurious state dependence effects for both homogeneous and heterogeneous logit models (rows 3 and 4, M1 vs. M2 and M3 vs. M5), but this is not the case for SBM choice set effects (rows 7 and 8, M1 vs. M2 and M3 vs. M5), at least in terms of RMSE(β). A parallel examination of selected means for -2MLL (Table 5) shows for both “All possible” and SBM consideration sets an improvement in fit due to spurious state dependence for the homogeneous logit models (rows 3, 4, 7 and 8, M1 vs. M2) but a *decrease* in fit for heterogeneous logit models (M3 vs. M5). Thus, there is the potential for spurious state dependence when there are underspecified consideration set effects, and the spurious effects may be beneficial or detrimental to parameter recovery and fit depending on the type of model (HB vs. homogeneous) and the type of consideration set effects. A similar analysis (Tables 4 and 5, rows 2, 4, 6, and 8, M2 vs. M5) shows spurious parameter heterogeneity when there are underspecified consideration set effects. The spurious effects are beneficial to parameter recovery in the case of "All possible" consideration set effects but detrimental in the case of SBM effects; fit always improves with spurious parameter heterogeneity.

Parameter Heterogeneity

Rows 9-24 of Table 4 examine the effects of underspecifying various forms of parameter heterogeneity. Rows 9-12 show that underspecifying normally-distributed parameter heterogeneity with variance .05 has minimal effects. In fact, it is not necessary to model such heterogeneity—when there are no consideration set effects (rows 10 and 12), homogeneous logit with GL (M2) is the preferred model. In contrast, underspecifying normal heterogeneity with variance .25 (rows 13-16) does become an issue. When there are no consideration set effects (rows 14 and 16), HB logit with GL (M5) is preferred to homogeneous logit with GL (M2).

One might expect models having mixtures of normal distributions to have an advantage for the fat tails condition (rows 17-20) and the bimodal condition (rows 21-24). To determine whether this is so, we can compare M8 (which should explain consideration set effects, state dependence, and unimodal normal heterogeneity) with M10 (which should explain all effects explained by M8 in addition to fat tails and bimodality). We see that there is very little difference between M8 and M10 in these conditions, indicating that there is little or no advantage to parameter recovery from using the mixture of normals specification over the usual normal distribution. However, there is evidence that the flexibility of the mixture of normals formulation could be useful in explaining other forms of unobserved heterogeneity. For example, compare the performance of M8 and M10 (which differ only according to the normal vs. mixture of normal formulation for heterogeneity) in the SBM choice set conditions with no parameter heterogeneity (rows 6 and 8). Despite the fact that there is no parameter heterogeneity in these conditions, the introduction of the mixture of normals component produces marked improvement of 35% and 27% in parameter recovery (but not fit—see Table 5). Though the mixture of normals component may not help us explain the exact nature of the choice process, the flexibility of the formulation does have the potential to explain other unobserved effects, which may result indirectly in improved parameter estimates. In fact, it is precisely this flexibility that results in M10 having lower overall parameter bias than M8.

Given parameter heterogeneity (rows 9-24), we can investigate the potential for spurious choice set effects and state dependence. When the only unobserved effect is parameter heterogeneity, adding a GL loyalty variable does not change the parameter estimates (rows 12, 16, 20, 24; compare M1 vs. M2, M3 vs. M5, M6 vs. M8, M7 vs. M9). This finding shows that we are unlikely to conclude that there is state dependence in the data when there is really parameter heterogeneity, consistent with the findings of AACJ. Likewise, when the only unobserved effect is parameter heterogeneity, adding a consideration set component to the model does not result in drastic changes in parameter estimates (rows 12, 16, 20, 24; compare M3 vs. M6, M5 vs. M8). Thus, the parameter estimates are unlikely to be affected when we model consideration effects while there is only parameter heterogeneity in the data. Unfortunately, the fit of a model with consideration set effects will be dramatically improved even when the only unobserved effect in the data is parameter heterogeneity (see Table 5, rows 12, 16, 20, 24, M3 vs. M6, M5 vs. M8), leading the analyst to believe that consideration set effects are present in the data. This improvement in fit occurs because the models with consideration set effects use the consumer's entire purchase history (both *past* and *future*) to narrow the consideration set for each purchase occasion, resulting in dramatic improvements in fit.

State Dependence

Rows 25-32 of Table 4 permit investigation of underspecified state dependence. When there is first-order state dependence (rows 25-28), the homogeneous logit model with GL (M2) is a strong model. The HB-estimated models do not perform as well as expected, even when the model is specified with a lagged dependent variable (M4). Fortunately, first-order state dependence does not appear to be common in scanner panel choice data.

As for higher-order GL-type state dependence in the data, adding the GL loyalty variable improves the performance of the homogeneous logit model (M1 vs. M2), though the improvement is due largely to spurious effects from underspecified consideration sets (row 31). Interestingly, when there are true higher-order state dependence effects in the data, the GL variable has little effect on parameter recovery for any of the heterogeneous models (compare M3 vs. M5, M6 vs. M8, and M7 vs. M9), and it is often *detrimental* to fit as well (see Table 5). Thus, there appears to be little rationale for including GL variables in heterogeneous logit models estimated with HB methods, whether or not there is real higher-order state dependence in the data. A lagged dependent variable works as well as the GL specification for the most part, and it is easier computationally since it does not require another layer of sampling in the Gibbs sampler.

Given state dependence in the data (rows 25-32), we can investigate the potential for spurious heterogeneity and consideration set effects. When the only unobserved effect in the data is state dependence, adding parameter heterogeneity to the model changes parameter estimates and fit (Tables 4 and 5, rows 28 and 32; compare M2 vs. M4 and M5). Importantly, we have the appearance of parameter heterogeneity even when there is only state dependence. Likewise, when the only unobserved effect in the data is state dependence, adding consideration set effects to the model changes parameter estimates and fit (rows 28 and 32; compare M5 vs. M8). Thus, we see evidence of parameter heterogeneity and consideration set effects even when there is only state dependence, and parameter estimates will suffer.

Finally, row 33 of Table 4 shows the RMSE(β) means when there are no unobserved effects in the data. All of the heterogeneous logit models indicate some degree of spurious effects. Also, in Table 5, row 33, compare the fit of the correct model M1 (homogeneous logit) with the best fitting model M8 (HB logit with consideration effects and GL): a 15% reduction in -2MLL occurs despite the fact that there are no unobserved effects in the data. Most of this spurious effect is due to parameter heterogeneity and consideration set effects—adding a lagged purchase variable or GL loyalty seems to have comparatively little effect.

Predictive Accuracy

Turning our attention now to the predictive accuracy of the models (Table 3), we see that, according to LOGL-V, M4 and M5 (HB logit with lagged dependent variable and GL loyalty, respectively) have the best overall predictive accuracy on a validation sample. Taken together, M4 and M5 offer about 6% improvement in validation fit over M2, M8, M9, and M10, about 19% improvement in fit over M6 and M7 and 27% improvement in fit over M1 and M3. Across data conditions, the only exception to this finding is for the "All possible" consideration sets condition, in which M8 (HB logit with consideration sets and GL) is best. M5 has the best forecasting performance even when there is no parameter heterogeneity in the data (primarily because it performs well in the "All possible" consideration sets scenario), which suggests that the validation sample log likelihood will not signal to the analyst that there are spurious effects. Table 6 shows selected means for LOGL-V. Note in particular that despite the spurious effects in row 33 (see Tables 4 and 5), the forecasting performance of the models is affected only when there is a consideration set component in the models (M6, M8, and M10). The impressive fit of the models with consideration set components does not carry over to the validation sample because sometimes consumers make purchases during the validation period that are outside the likely consideration sets identified from the estimation sample.

Interestingly, three different models are shown to be best on the three different criteria RMSE(β), -2MLL, and LOGL-V, which suggests that it will be difficult to pick the model that best recovers the true parameters using fit and prediction measures. Table 7 shows the correlations between RMSE(β), -2MLL, and LOGL-V for each model type. Ideally, we would like to see strong positive correlations between RMSE(β) and -2MLL (better fit implies lower parameter estimation error), strong negative correlations between RMSE(β) and LOGL-V (better forecasting accuracy implies lower parameter estimation error) and strong negative correlations between -2MLL and LOGL-V. Unfortunately, except for the homogeneous logit model (M1), we do not see such a pattern. Though several models show strong negative correlations between -2MLL and LOGL-V, none of the HB models shows a strong positive correlation between -2MLL and RMSE(β), and the correlations between LOGL-V and RMSE(β) are disappointing as well. Thus, much work remains to be done in developing model selection and evaluation criteria that are useful for HB logit models.

Summary and Conclusions

The main findings of this study are summarized in the following paragraphs.

1. Underspecifying consideration sets results in major parameter bias for HB logit models with individual level parameters, as it does for homogeneous logit models. Underspecified consideration set effects can result in spurious state dependence, which may be beneficial for parameter estimates and fit (in the case of homogeneous logit) or detrimental (in the case of HB logit), depending on the type of consideration set effects. Thus, the consequences of these spurious effects are different for HB models vs. homogeneous logit models (AACJ did not consider HB models). Underspecified consideration set effects also result in spurious parameter heterogeneity (which can also be beneficial or detrimental, depending on the type of consideration set effects) but there is also evidence that spurious parameter heterogeneity can occur even when there are no unobserved effects at all.

2. When true parameter heterogeneity is normally distributed and small (e.g., variance of .05), it is not necessary to model it, though there are benefits to modeling more extensive heterogeneity (e.g., variance of .25). In general, models with mixtures of normal distributions do not produce improved parameter estimates when the assumption of normally-distributed coefficients is violated (e.g., fat-tailed distributions or bimodal distributions). Though the mixture of normals formulation does not help us explain the exact nature of preferences and the choice process, the flexibility of the formulation does have the potential to explain other unobserved effects, which may result indirectly in improved parameter estimates. Consistent with the findings of AACJ, we do not observe major changes to the parameter estimates when state dependence or consideration set effects are wrongly used to model parameter heterogeneity, though the fit statistic does suggest spurious consideration set effects. Models with consideration set effects will usually fit better than those without because they use the consumer's entire purchase history (past and future) to narrow the consideration set for each purchase occasion, resulting in a dramatic improvement in fit, regardless of whether there are true consideration set effects in the data. As a result, there is evidence that spurious consideration set effects can occur even when there are no unobserved effects at all.
3. When there is true higher-order state dependence in choice data, adding a GL loyalty variable to an HB model with individual level parameters makes very little difference in parameter estimates and often hurts fit, which is in stark contrast to the findings for homogeneous logit models (cf. the findings of AACJ). An individual level HB logit model with lagged dependent variable has equally accurate parameter estimates,

slightly better fit, and equally accurate predictions compared to an individual level HB logit model with GL variable. Including a lagged dependent variable rather than a GL loyalty variable in an HB model is preferable from a computational standpoint as well. We see evidence of parameter heterogeneity and consideration set effects when there is underspecified state dependence, but as mentioned above, we can see these effects even when there are no unobserved effects at all in the data.

4. The validation sample log likelihood will not always identify spurious effects. For example, an individual-level HB logit model with GL loyalty has the best forecasting performance even when there is no parameter heterogeneity in the data, primarily because it does a good job of explaining "all possible" consideration set effects. LOGL-V will rarely identify a model with consideration set effects as being best because households sometimes make a "surprise" brand choice outside the likely consideration sets during the validation period. Thus, it is difficult to rely on LOGL-V for information on the validity of the consideration set component of an individual level HB model.
5. In general, fit and forecasting accuracy provide no consistent information for identifying models with the lowest parameter bias.
6. A fully-specified model with mixture of normals parameter heterogeneity, consideration set effects, and GL loyalty variable produced the best parameter estimates overall (M10). A model with normally-distributed parameter heterogeneity, consideration set effects, and GL loyalty produced the best fit to the estimation data (M8). A model with normally-distributed parameter heterogeneity and either a lagged dependent variable (M4) or GL loyalty variable (M5) produced the best

predictions for a validation sample. M10, M8, and M5 have not been estimated previously in the choice modeling literature.

These results suggest that model selection should be based on managerial objectives for model building such as marketing mix planning, market segmentation, or forecasting. If the model is being developed for marketing mix planning or market segmentation purposes, it will be important for marketing modelers to assess parameter recovery of different models, which is rarely if ever assessed today. It is risky to rely on fit or predictive validation statistics because they are found to be uncorrelated with parameter estimation error. The flexibility afforded by a fully-specified model containing consideration set effects, loyalty or some form of purchase event feedback, and mixtures of normals heterogeneity is likely to be most effective when parameter recovery is important, while a simpler model with state dependence but without consideration set effects or mixture of normals heterogeneity is likely to offer the best forecasts.

Though the potential benefits of using HB models are clear, this study highlights some areas that are in need of further research. First, more research is needed on methods of incorporating consideration set effects into HB logit models. As well as the consideration set model based on the Chiang, Chib, and Narasimhan (1999) study performed with data that matched the model specification, it did not capture dynamic consideration set formation (Siddarth, Bucklin, and Morrison 1995). Also, since the model is based on the power set of available brands, it is not applicable to markets with more than 8-10 brands due to the computational burden involved. Given the importance of modeling consideration set effects in HB logit models, researchers need to continue development of models capable of explaining these effects. For an example of pioneering new work in this area, see Gilbride and Allenby (2003).

Second, work remains to be done on the development of model selection criteria for HB logit models. Though studies by Allenby and Rossi (1999) and Rossi and Allenby (2003) do address misspecification testing in HB models, our study suggests that marginal log likelihood and validation sample log likelihood measures have very low correlations with parameter bias, making the selection of a low-bias model difficult. A recent study by Andrews and Currim (2003) examines the issue of model selection criteria for finite mixture logit models but does not address criteria for HB logit models.

Finally, researchers should consider carefully the finding that HB logit models can produce good parameter estimates, fit, and predictions yet provide inaccurate descriptions of the choice process used by consumers. Sometimes spurious effects can actually improve parameter estimates, fit, and predictions, even though the process explanation is not accurate. One could make the case that logit models formulated and estimated from a Bayesian perspective are more suitable for certain kinds of research (e.g., forecasting, marketing mix planning) and less suitable for others (e.g., theory testing, process modeling). Even so, with the expanded availability of micro-level marketplace data and continuing developments in computational methods, the Bayesian approach to choice modeling has the potential to improve decision making for managers of frequently purchased packaged goods.

References

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, 37 (November), 410-426.

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

__________ and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," Journal of Econometrics, 89 (March/April), 57-78.

Andrews, Rick L., Andrew Ainslie, and Imran S. Currim (2002), "An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity," Journal of Marketing Research, 39 (November), 479-487.

__________, Asim Ansari, and Imran S. Currim (2002), "Hierarchical Bayes vs. Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery," Journal of Marketing Research, 39 (February), 87-98.

__________ and Imran S. Currim (2003), "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models," Journal of Marketing Research, 40 (May), 235-243.

Chiang, Jeongwen, Siddhartha Chib, and Chakravarthi Narasimhan (1999), "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," Journal of Econometrics, 89 (1-2), 223-248.

Fader, Peter S. and James M. Lattin (1993), "Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purcase Behavior," Marketing Science, 12 (Summer), 304-17.

Gilbride, Timothy and Greg Allenby (2003), "A Choice Model with Conjunctive, Disjunctive and Compensatory Screening Rules," working paper, OSU

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

Newton, M.A. and Raftery, A.E. (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," Journal of the Royal Statistical Society, Series (B), 56, 43-48.

Raftery, Adrian E. (1996), “Hypothesis Testing and Model Selection,” in Markov Chain Monte Carlo in Practice, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds., London: Chapman & Hall, 163-188.

Rossi, Peter E. and Greg M. Allenby (2003), “Bayesian Statistics and Marketing,” Marketing Science, 22 (Summer), 304-328.

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, 15 (3), 280-299.

Schwartz, G. (1978) “Estimating the Dimension of a Model,” Annals of Statistics, 6, 461-464.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison (1995), "Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data," Journal of Marketing Research, 32 (August), 255-66.

Vriens, Marco, Michel Wedel, and Tom Wilms (1996), “Metric Conjoint Segmentation Methods: A Monte Carlo Comparison,” Journal of Marketing Research, 33 (February), 73-85.

Wedel, Michel, Wagner Kamakura, Neeraj Arora, Albert Bemmaor, Jeongwen Chiang, Terry Elrod, Rich Johnson, Peter Lenk, Scott Neslin, and Carsten Stig Poulsen (1999), “Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modeling,” Marketing Letters, 10 (3), 219-232.

Table 1. Summary of Results for Model Parameter Recovery (RMSE(β)).

Factor	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**
# Purchases										
5	0.85	0.59	0.60	0.61	0.60	0.61	0.70	0.69	0.63	**0.53**
10	0.85	0.56	0.56	0.53	0.53	0.52	0.60	0.47	0.58	**0.45**
15	0.83	0.52	0.53	0.45	0.46	0.50	0.57	**0.40**	0.51	0.41
# Households										
100	0.85	0.56	0.59	0.55	0.54	0.58	0.69	0.57	0.60	**0.51**
200	0.84	0.55	0.55	0.53	0.53	0.53	0.61	0.49	0.57	**0.44**
300	0.84	0.56	0.55	0.52	0.52	0.52	0.58	0.49	0.56	**0.44**
Consideration Sets										
None	0.55	**0.37**	0.46	0.39	0.39	0.49	0.51	0.44	0.40	0.43
All possible	1.30	0.83	0.65	0.64	0.68	0.56	0.80	0.51	0.83	**0.47**
SBM	0.67	**0.47**	0.58	0.56	0.53	0.57	0.57	0.61	0.49	0.50
Heterogeneity										
None	0.67	0.36	0.44	0.41	0.41	0.43	0.47	0.41	0.41	**0.29**
Normal, .05 variance	0.77	0.47	0.49	0.46	0.46	0.46	0.56	0.45	0.49	**0.38**
Normal, .25 variance	0.94	0.68	0.66	0.61	0.61	0.63	0.74	0.60	0.67	**0.57**
Mixture, fat tails	0.90	0.63	0.61	0.58	0.59	0.59	0.69	0.57	0.64	**0.53**
Mixture, bimodal	0.93	0.66	0.62	0.59	0.59	0.60	0.69	0.56	0.65	**0.55**
State Dependence										
None	0.68	0.58	0.49	0.52	0.54	0.48	0.54	0.48	0.59	**0.45**
1st order	1.09	0.55	0.68	0.55	0.54	0.66	0.80	0.58	0.56	**0.49**
Guadagni-Little	0.76	0.55	0.51	0.53	0.52	0.48	0.54	0.50	0.57	**0.45**
Means	0.84^{6}	0.56^{4}	0.56^{4}	$0.53^{2,3}$	$0.53^{2,3}$	0.54^{3}	0.63^{5}	0.52^{2}	0.57^{4}	**0.46^{1}**

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 2. Summary of Results for Model Fit (-2MLL).

Factor	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
# Purchases										
5	2.33	1.74	1.29	1.48	1.54	1.04	1.74	**0.99**	1.69	1.04
10	2.31	1.66	1.39	1.38	1.45	1.28	1.60	**1.16**	1.60	1.20
15	2.29	1.63	1.46	1.37	1.41	1.39	1.60	**1.24**	1.55	1.28
# Households										
100	2.32	1.70	1.40	1.43	1.49	1.26	1.73	**1.14**	1.63	1.19
200	2.31	1.67	1.37	1.40	1.46	1.23	1.63	**1.12**	1.61	1.17
300	2.30	1.67	1.37	1.40	1.46	1.23	1.58	**1.13**	1.60	1.17
Consideration Sets										
None	2.00	1.68	1.56	1.54	1.56	1.43	1.80	**1.30**	1.63	1.35
All possible	2.76	1.68	1.10	1.20	1.31	0.96	1.37	**0.88**	1.57	0.92
SBM	2.17	1.68	1.48	1.49	1.54	1.33	1.77	**1.20**	1.63	1.26
Heterogeneity										
None	2.24	1.62	1.37	1.38	1.44	1.23	1.62	**1.10**	1.56	1.13
Normal, .05 variance	2.28	1.65	1.38	1.41	1.46	1.23	1.63	**1.12**	1.59	1.16
Normal, .25 variance	2.34	1.71	1.39	1.43	1.49	1.25	1.67	**1.15**	1.65	1.20
Mixture, fat tails	2.34	1.70	1.39	1.44	1.49	1.26	1.66	**1.15**	1.64	1.20
Mixture, bimodal	2.35	1.70	1.37	1.40	1.46	1.23	1.65	**1.13**	1.63	1.19
State Dependence										
None	2.19	1.93	1.50	1.60	1.66	**1.33**	1.75	**1.33**	1.86	1.38
1st order	2.51	1.38	1.36	1.20	1.20	1.24	1.61	**0.90**	1.32	0.94
Guadagni-Little	2.23	1.72	1.29	1.43	1.54	**1.15**	1.58	1.16	1.67	1.21
Means	2.31^{10}	1.68^{9}	1.38^{4}	1.41^{5}	1.47^{6}	1.24^{3}	1.65^{8}	**1.13^{1}**	1.61^{7}	1.18^{2}

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 3. Summary of Results for Validation of Model Predictive Accuracy (LOGL-V).

Factor	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
# Purchases										
5	-1.13	-0.84	-1.09	-0.82	**-0.81**	-0.97	-1.06	-0.87	-0.84	-0.87
10	-1.13	-0.83	-0.98	**-0.77**	**-0.77**	-0.93	-0.96	-0.81	-0.81	-0.82
15	-1.12	-0.84	-0.92	-0.75	**-0.74**	-0.91	-0.92	-0.78	-0.79	-0.79
# Households										
100	-1.13	-0.84	-1.01	-0.79	**-0.78**	-0.94	-0.99	-0.82	-0.83	-0.83
200	-1.13	-0.83	-0.99	-0.78	**-0.77**	-0.93	-0.97	-0.82	-0.81	-0.82
300	-1.13	-0.83	-0.99	-0.78	**-0.77**	-0.93	-0.98	-0.81	-0.80	-0.82
Consideration Sets										
None	-0.99	-0.82	-1.01	-0.82	**-0.81**	-1.07	-1.00	-0.95	-0.82	-0.96
All possible	-1.37	-0.86	-0.91	-0.70	-0.69	-0.70	-0.92	**-0.60**	-0.79	-0.61
SBM	-1.03	-0.82	-1.06	-0.82	**-0.81**	-1.03	-1.03	-0.90	-0.82	-0.91
Heterogeneity										
None	-1.09	-0.81	-0.99	-0.77	**-0.76**	-0.93	-0.96	-0.80	-0.79	-0.80
Normal, .05 variance	-1.11	-0.82	-0.99	-0.77	**-0.76**	-0.92	-0.97	-0.81	-0.80	-0.81
Normal, .25 variance	-1.15	-0.85	-1.01	-0.80	**-0.79**	-0.95	-0.99	-0.84	-0.83	-0.85
Mixture, fat tails	-1.15	-0.85	-1.00	-0.79	**-0.78**	-0.94	-1.00	-0.83	-0.82	-0.84
Mixture, bimodal	-1.15	-0.85	-1.00	-0.78	**-0.77**	-0.93	-0.99	-0.82	-0.82	-0.83
State Dependence										
None	-1.07	-0.97	**-0.88**	**-0.88**	**-0.88**	-0.93	-0.91	-0.92	-0.93	-0.93
1st order	-1.23	-0.67	-1.25	**-0.65**	**-0.65**	-1.00	-1.16	-0.69	-0.66	-0.70
Guadagni-Little	-1.09	-0.86	-0.87	-0.82	**-0.79**	-0.88	-0.88	-0.84	-0.84	-0.85
Means	-1.13^{8}	-0.84^{4}	-1.00^{7}	**-0.78^{1}**	**-0.77^{1}**	-0.94^{5}	-0.98^{6}	-0.82^{2,3}	-0.81^{2}	-0.83^{3,4}

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 4. Selected Cell Means for Model Parameter Recovery (RMSE(β)).

Row	CS	Het	SD	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
1	All			1.30	0.83	0.65	0.64	0.68	0.56	0.80	0.51	0.83	**0.47**
2	All	N		1.21	0.70	0.53	0.50	0.55	0.46	0.65	0.38	0.71	**0.28**
3	All		N	1.17	0.89	0.57	0.63	0.72	0.47	0.73	0.47	0.89	**0.45**
4	All	N	N	1.15	0.81	0.45	0.52	0.66	0.39	0.56	0.36	0.80	**0.26**
5	SBM			0.67	**0.47**	0.58	0.56	0.53	0.57	0.57	0.61	0.49	0.50
6	SBM	N		0.50	**0.28**	0.45	0.48	0.44	0.45	0.41	0.52	0.33	0.34
7	SBM		N	0.49	**0.48**	0.54	0.55	0.51	0.54	0.49	0.53	0.50	**0.48**
8	SBM	N	N	**0.27**	**0.27**	0.44	0.47	0.40	0.46	0.31	0.44	0.32	0.32
9		.05		0.77	0.47	0.49	0.46	0.46	0.46	0.56	0.45	0.49	**0.38**
10	N	.05		0.43	**0.25**	0.37	0.30	0.30	0.41	0.41	0.38	0.29	0.33
11		.05	N	0.60	0.50	0.43	0.44	0.48	0.40	0.46	0.41	0.52	**0.35**
12	N	.05	N	**0.25**	**0.25**	0.32	0.30	0.33	0.37	0.29	0.38	0.29	0.33
13		.25		0.94	0.68	0.66	0.61	0.61	0.63	0.74	0.60	0.67	**0.57**
14	N	.25		0.70	0.54	0.57	**0.49**	**0.49**	0.58	0.65	0.53	0.54	0.54
15		.25	N	0.77	0.69	0.58	0.59	0.61	**0.55**	0.65	**0.55**	0.69	0.56
16	N	.25	N	0.54	0.54	**0.49**	**0.49**	0.50	0.52	0.55	0.51	0.53	0.53
17		Fat		0.90	0.63	0.61	0.58	0.59	0.59	0.69	0.57	0.64	**0.53**
18	N	Fat		0.65	0.47	0.52	**0.45**	**0.45**	0.54	0.58	0.49	0.48	0.49
19		Fat	N	0.74	0.64	0.53	0.57	0.58	0.53	0.61	0.54	0.66	**0.52**
20	N	Fat	N	0.47	0.47	**0.44**	0.45	0.45	0.47	0.49	0.49	0.49	0.49
21		Bi		0.93	0.66	0.62	0.59	0.59	0.60	0.69	0.56	0.65	**0.55**
22	N	Bi		0.68	0.52	0.52	**0.45**	**0.45**	0.55	0.59	0.50	0.50	0.52
23		Bi	N	0.77	0.68	0.56	0.58	0.59	0.53	0.63	**0.52**	0.66	0.54
24	N	Bi	N	0.55	0.54	0.43	**0.42**	0.43	0.47	0.51	0.45	0.49	0.50
25			1st	1.09	0.55	0.68	0.55	0.54	0.66	0.80	0.58	0.56	**0.49**
26	N		1st	0.85	**0.37**	0.60	0.40	0.39	0.62	0.71	0.48	0.41	0.44
27		N	1st	0.97	**0.33**	0.55	0.43	0.41	0.56	0.68	0.47	0.40	**0.33**
28	N	N	1st	0.73	**0.07**	0.48	0.27	0.24	0.53	0.60	0.37	0.21	0.25
29			GL	0.76	0.55	0.51	0.53	0.52	0.48	0.54	0.50	0.57	**0.45**
30	N		GL	0.43	**0.38**	0.40	0.39	**0.38**	0.43	0.42	0.43	0.40	0.41
31		N	GL	0.55	0.35	0.40	0.39	0.38	0.35	0.36	0.39	0.41	**0.26**
32	N	N	GL	0.12	**0.10**	0.30	0.26	0.24	0.33	0.19	0.31	0.20	0.22
33	N	N	N	**0.08**	**0.08**	0.21	0.25	0.24	0.25	0.21	0.29	0.17	0.25

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 5. Selected Cell Means for Model Fit (-2MLL).

Row	CS	Het	SD	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
1	All			2.76	1.68	1.10	1.20	1.31	0.96	1.37	**0.88**	1.57	0.92
2	All	N		2.73	1.66	1.08	1.18	1.30	0.94	1.34	**0.86**	1.55	0.87
3	All		N	2.70	2.00	1.23	1.36	1.48	**1.03**	1.55	**1.03**	1.88	1.07
4	All	N	N	2.67	1.94	1.17	1.29	1.42	**0.96**	1.46	**0.96**	1.78	0.98
5	SBM			2.17	1.68	1.48	1.49	1.54	1.33	1.77	**1.20**	1.63	1.26
6	SBM	N		2.12	1.63	1.50	1.48	1.52	1.34	1.78	**1.19**	1.59	1.23
7	SBM		N	2.02	1.94	1.63	1.75	1.80	**1.45**	1.91	**1.45**	1.90	1.51
8	SBM	N	N	1.93	1.87	1.60	1.70	1.75	**1.42**	1.87	**1.42**	1.83	1.46
9		.05		2.28	1.65	1.38	1.41	1.46	1.23	1.63	**1.12**	1.59	1.16
10	N	.05		1.96	1.64	1.58	1.55	1.56	1.44	1.80	1.31	1.61	1.34
11		.05	N	2.13	1.88	1.48	1.58	1.64	**1.30**	1.71	**1.30**	1.80	1.34
12	N	.05	N	1.74	1.75	1.59	1.66	1.67	**1.46**	1.70	**1.46**	1.71	1.48
13		.25		2.34	1.71	1.39	1.43	1.49	1.25	1.67	**1.15**	1.65	1.20
14	N	.25		2.07	1.74	1.59	1.59	1.60	1.44	1.83	**1.33**	1.70	1.38
15		.25	N	2.21	1.97	1.53	1.63	1.69	**1.36**	1.79	**1.36**	1.90	1.42
16	N	.25	N	1.90	1.91	1.66	1.77	1.76	**1.53**	1.85	**1.53**	1.86	1.58
17		Fat		2.34	1.70	1.39	1.44	1.49	1.26	1.66	**1.15**	1.64	1.20
18	N	Fat		2.05	1.73	1.58	1.58	1.60	1.45	1.84	**1.34**	1.69	1.40
19		Fat	N	2.23	1.96	1.52	1.64	1.71	1.36	1.80	**1.35**	1.90	1.41
20	N	Fat	N	1.91	1.91	1.68	1.75	1.78	**1.55**	1.86	**1.55**	1.87	1.60
21		Bi		2.35	1.70	1.37	1.40	1.46	1.23	1.65	**1.13**	1.63	1.19
22	N	Bi		2.04	1.69	1.54	1.51	1.53	1.40	1.76	**1.28**	1.63	1.34
23		Bi	N	2.24	1.97	1.51	1.62	1.67	**1.34**	1.77	**1.34**	1.90	1.41
24	N	Bi	N	1.90	1.89	1.62	1.70	1.71	**1.48**	1.81	1.49	1.82	1.56
25			1st	2.51	1.38	1.36	1.20	1.20	1.24	1.61	**0.90**	1.32	0.94
26	N		1st	2.27	1.45	1.59	1.33	1.33	1.45	1.87	1.07	1.41	1.11
27		N	1st	2.46	1.33	1.36	1.17	1.18	1.24	1.61	**0.88**	1.27	0.90
28	N	N	1st	2.15	1.36	1.56	1.26	1.27	1.42	1.86	**1.01**	1.31	1.04
29			GL	2.23	1.72	1.29	1.43	1.54	**1.15**	1.58	1.16	1.67	1.21
30	N		GL	1.89	1.74	1.46	1.58	1.62	**1.33**	1.74	1.34	1.70	1.39
31		N	GL	2.16	1.69	1.28	1.41	1.52	1.15	1.57	**1.14**	1.63	1.18
32	N	N	GL	1.75	1.64	1.41	1.52	1.54	1.29	1.68	**1.28**	1.59	1.31
33	N	N	N	1.73	1.74	1.59	1.65	1.67	1.48	1.70	**1.47**	1.71	1.50

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 6. Selected Cell Means for Validation of Model Predictive Accuracy (LOGL-V).

Row	CS	Het	SD	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
1	All			-1.37	-0.86	-0.91	-0.70	-0.69	-0.70	-0.92	**-0.60**	-0.79	-0.61
2	All	N		-1.35	-0.85	-0.91	-0.69	-0.68	-0.70	-0.91	**-0.59**	-0.77	**-0.59**
3	All		N	-1.34	-1.06	-0.79	-0.79	-0.80	-0.68	-0.86	**-0.67**	-0.95	-0.68
4	All	N	N	-1.32	-1.02	-0.75	-0.75	-0.75	**-0.62**	-0.79	**-0.62**	-0.87	**-0.62**
5	SBM			-1.03	-0.82	-1.06	-0.82	**-0.81**	-1.03	-1.03	-0.90	-0.82	-0.91
6	SBM	N		-1.00	**-0.80**	-1.06	-0.81	**-0.80**	-1.04	-1.02	-0.90	-0.81	-0.90
7	SBM		N	-0.97	-0.96	-0.95	-0.95	**-0.94**	-1.06	-0.96	-1.05	-0.95	-1.06
8	SBM	N	N	-0.92	**-0.92**	-0.93	-0.93	**-0.92**	-1.05	-0.93	-1.04	-0.92	-1.04
9		.05		-1.11	-0.82	-0.99	-0.77	**-0.76**	-0.92	-0.97	-0.81	-0.80	-0.81
10	N	.05		-0.96	**-0.80**	-1.02	-0.81	-0.81	-1.06	-0.99	-0.94	-0.81	-0.94
11		.05	N	-1.05	-0.96	**-0.86**	**-0.86**	-0.87	-0.91	-0.89	-0.90	-0.91	-0.90
12	N	.05	N	-0.86	**-0.86**	-0.87	-0.87	-0.87	-1.02	-0.87	-1.02	-0.87	-1.02
13		.25		-1.15	-0.85	-1.01	-0.80	**-0.79**	-0.95	-0.99	-0.84	-0.83	-0.85
14	N	.25		-1.03	-0.86	-1.03	-0.85	**-0.84**	-1.10	-1.02	-0.98	-0.86	-0.99
15		.25	N	-1.08	-0.99	-0.91	**-0.90**	**-0.90**	-0.95	-0.92	-0.94	-0.96	-0.95
16	N	.25	N	-0.94	-0.94	**-0.92**	-0.93	-0.93	-1.07	-0.94	-1.07	-0.94	-1.08
17		Fat		-1.15	-0.85	-1.00	-0.79	**-0.78**	-0.94	-1.00	-0.83	-0.82	-0.84
18	N	Fat		-1.02	-0.85	-1.03	-0.85	**-0.84**	-1.09	-1.03	-0.97	-0.85	-0.99
19		Fat	N	-1.10	-1.00	**-0.90**	**-0.90**	**-0.90**	-0.95	-0.94	-0.95	-0.97	-0.96
20	N	Fat	N	-0.95	-0.95	**-0.94**	**-0.94**	**-0.94**	-1.09	-0.95	-1.09	-0.95	-1.10
21		Bi		-1.15	-0.85	-1.00	-0.78	**-0.77**	-0.93	-0.99	-0.82	-0.82	-0.83
22	N	Bi		-1.00	-0.82	-1.00	-0.81	**-0.79**	-1.05	-0.99	-0.93	-0.81	-0.95
23		Bi	N	-1.09	-0.98	**-0.88**	**-0.88**	**-0.88**	-0.93	-0.92	-0.92	-0.94	-0.94
24	N	Bi	N	-0.90	-0.89	**-0.86**	**-0.86**	**-0.86**	-1.01	-0.88	-1.01	-0.88	-1.03
25			1st	-1.23	-0.67	-1.25	-0.65	**-0.65**	-1.00	-1.16	-0.69	-0.66	-0.70
26	N		1st	-1.13	**-0.71**	-1.25	**-0.71**	**-0.71**	-1.15	-1.18	-0.83	**-0.71**	-0.84
27		N	1st	-1.20	-0.65	-1.24	**-0.63**	**-0.63**	-1.01	-1.15	-0.68	-0.64	-0.69
28	N	N	1st	-1.08	-0.66	-1.22	-0.68	-0.68	-1.13	-1.14	-0.81	**-0.67**	-0.80
29			GL	-1.09	-0.86	-0.87	-0.82	**-0.79**	-0.88	-0.88	-0.84	-0.84	-0.85
30	N		GL	-0.94	-0.86	-0.90	-0.87	**-0.84**	-1.01	-0.91	-0.98	-0.86	-0.99
31		N	GL	-1.05	-0.85	-0.87	-0.82	**-0.79**	-0.88	-0.88	-0.84	-0.83	-0.84
32	N	N	GL	-0.88	**-0.82**	-0.88	-0.84	**-0.82**	-0.98	-0.88	-0.95	**-0.82**	-0.95
33	N	N	N	**-0.86**	**-0.86**	**-0.86**	-0.87	-0.87	-1.00	**-0.86**	-1.00	**-0.86**	-0.99

M1: Homogeneous logit, zero-order
M2: Homogeneous logit, Guadagni-Little (GL) loyalty
M3: HB logit
M4: HB logit with lagged purchase variable
M5: HB logit with GL loyalty
M6: HB logit with consideration sets
M7: HB logit with mixture of normals
M8: HB logit with consideration sets and GL
M9: HB logit with mixture of normals and GL
M10: HB logit with consideration sets, mixture of normals, and GL

Table 7. Correlations between Model Parameter Recovery (RMSE(β)), Fit (-2MLL), and Validation of Predictive Accuracy (LOGL-V), by Model Type

M1: Homogeneous logit, zero-order

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	0.90	1.00	
LOGL-V	-0.88	-0.98	1.00

M2: Homogeneous logit, Guadagni-Little (GL) loyalty

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	0.14	1.00	
LOGL-V	-0.20	-0.93	1.00

M3: HB logit

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.33	1.00	
LOGL-V	-0.30	-0.11	1.00

M4: HB logit with lagged purchase variable

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.17	1.00	
LOGL-V	0.12	-0.93	1.00

M5: HB logit with GL

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.62	1.00	
LOGL-V	0.16	-0.42	1.00

M6: HB logit with consideration sets

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	0.11	1.00	
LOGL-V	-0.13	-0.72	1.00

M7: HB logit with mixture of normals

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.46	1.00	
LOGL-V	-0.35	-0.02	1.00

M8: HB logit with consideration sets and GL

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.37	1.00	
LOGL-V	-0.01	-0.78	1.00

M9: HB logit with mixture of normals and GL

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	0.08	1.00	
LOGL-V	-0.06	-0.95	1.00

M10: HB logit with consideration sets, mixture of normals, and GL

	RMSE(β)	-2MLL	LOGL-V
RMSE(β)	1.00		
-2MLL	-0.17	1.00	
LOGL-V	0.00	-0.81	1.00