The authors present a Bayesian approach to the estimation of household parameters. Applied to the standard logit model, the procedure produces household-level estimates of all model parameters, enabling researchers to identify differences in household reaction to all variables in the marketing mix. Simulated data are used to study the small-sample performance of the estimator. The estimator can be easily implemented with standard algorithms used to maximize likelihood functions. In application to tuna scanner panel data, strong evidence of heterogeneity in price, display, and feature response (slope) parameters is detected. Approaches that fail to take into account slope heterogeneity are shown to underestimate the value of feature advertising and in-store displays in this dataset. In addition, the household price sensitivity estimates are strongly related to coupon usage and demonstrate how the estimates can be used to implement a targeted household drop.

A Bayesian Approach to Estimating Household Parameters

A fundamental assumption of marketing is that consumers differ in their preferences and reactions to promotional variables such as price and advertising. Researchers have proposed a wide variety of models of household behavior as a basis for developing and evaluating marketing policies. At the same time, a wealth of information on household purchase behavior has become available in scanner panel datasets. However, attempts to use household-level models with panel data have been frustrated by the lack of statistical procedures that are flexible enough to generate household-level estimates for all model parameters. We introduce a new approach to estimating household parameters that yields estimates of household preference and reaction to promotional variables.

Our household-level estimates can be used to address two important issues. First, household estimates can be used to tailor marketing strategies to specific households. We give an example in which direct mail coupon drops targeted to price-sensitive households are used instead of traditional blanket mailings. Second, market-wide response to promotional variables can be estimated by aggregating the estimated responses across households. Estimates of aggregate market response can be severely biased unless adjustments are made for differences among households. In our analysis of scanner panel data, estimates of market response to in-store display and newspaper advertising double in size when adjustments are made for the distribution of heterogeneity.

The challenge facing the modeler using scanner panel data is to devise a flexible estimation procedure that incorporates heterogeneity of a general form while simultaneously reckoning with the small amount of data per household. Typically, no more than 25 or so purchases a year are observed for each panel household, even though there might be more than 500 households in the panel. As we discuss in detail subsequently, most current procedures meet this modeling challenge by assuming that some of the model parameters are constant across all households (Guadagni and Little 1983) or at least some subset of households (Currim 1981; Kamakura and Russell 1989). An alternative approach is to assume that household parameters come from a common distribution. This approach, referred to as a random effects specification, estimates the distribution of parameters over

*Peter E. Rossi is Associate Professor, Graduate School of Business, University of Chicago. Greg M. Allenby is Assistant Professor, College of Business, The Ohio State University.

The authors appreciate helpful comments by Bob Blattberg, Peter Dickson, Kris Helsen, Jim Ginter, and Mary Sullivan. In addition, the editor and anonymous JMR reviewers made invaluable suggestions for improvement of drafts of the article. The authors are listed in reverse alphabetical order.
the population but does not directly produce household parameter estimates. A random effects model for the intercepts of a logit model is investigated by Chintagunta, Jain, and Vilcassim (1991). Our analysis of scanner data suggests that it is very important to allow for different slopes as well as intercepts across the population of households.

All current approaches to modeling heterogeneity in short panels rely on some sort of a priori restrictions on the extent of variation in the parameters across households. For example, Kamakura and Russell (1989) assume that households within a segment share common parameter values. We solve the problem of scarcity of household information by using prior information about the likely range of choice probabilities and parameter values. Our approach incorporates prior information through a Bayesian method that yields different parameter estimates for each household. An interpretable and flexible reference prior is used to form the household Bayes estimates. Moreover, our Bayesian household estimates can be obtained by simple modifications of current maximum likelihood algorithms.

In the next section, we review and critique present procedures in more detail. The proposed Bayesian procedures are then introduced and evaluated with both simulated and actual scanner panel data. The simulation experiments serve to assess the sampling properties of the proposed procedures and to compare them with alternatives. We then use an ERIM scanner panel dataset from A. C. Nielsen to illustrate the procedure and provide new evidence on the bias induced by slope heterogeneity and the relation between price sensitivity and coupon usage. Finally, we examine the predictive performance of the procedure and offer concluding remarks.

**REVIEW AND CRITIQUE OF PRESENT PROCEDURES**

Approaches to estimating parameters in models of consumer behavior can be grouped into two broad classes, (1) those that estimate parameters for each household (i.e., a fixed effect model) and (2) those that assume the household parameters are distributed according to a probability distribution and estimate the parameters of that distribution (random effects models).

The Random Effects Approach

A natural and conceptually appealing approach to modeling heterogeneity across households is to assume that the parameters vary according to a probability distribution. This general approach is called a random effects model. The advantage of the random effects model comes from the radical reduction in the number of parameters achieved by assuming a distribution of heterogeneity. Suppose the basic statistical model has \( k \) parameters. If we start with \( N \) households, the random effects approach replaces the \( N \times k \) household-level parameters with a smaller number of parameters that describe the heterogeneity (or mixture) distribution.

Key to the success of the random effects approach is the proper choice of the heterogeneity distribution. There is a tradeoff between flexibility and parsimony. The analyst would like to specify a very general form for the distribution to avoid specification errors that might arise from assuming an incorrect form. An alternative approach advocated by Kamakura and Russell (1989) is the finite mixture model, which approximates the continuous distribution of heterogeneity by a discrete distribution. That model assumes a fixed number of homogeneous segments and estimates segment parameters and segment sizes simultaneously. Our view is that the question of whether the distribution of heterogeneity is discrete (with a small number of mass points) or continuous should be answered from the data. Our procedures for estimating household parameters, which do not assume a fixed number of segments or any particular form for the heterogeneity distribution, can shed light on this issue.

Perhaps the most severe limitation of all random effects models is that they are not well suited for the estimation of household parameters. For some marketing decisions such as a chainwide pricing decision, knowledge of the distribution of heterogeneity is sufficient for correct prediction of aggregate market response. However, this knowledge is not adequate for designing customized marketing instruments, which are increasingly being used to target specific households. Examples include direct mail and point-of-purchase couponing. Given the technological revolution in the collection of purchasing data from a wide variety of sources, household-specific marketing instruments should become even more common in the near future.

**Fixed Effects Models**

Difficulties in estimating fixed effects models at the household level have prevented the investigation of heterogeneity in all model parameters. Heterogeneity is typically modeled by including demographic variables or variables that measure past buying behavior as covariates (Allenby and Rossi 1991a; Guadagni and Little 1983). These variables, which alter the intercepts for each household, are useful in studying the heterogeneity of consumer preference for different brands. However, there is evidence that the standard set of household demographic data is inadequate to capture heterogeneity. In addition, it is implausible that heterogeneity is limited to the intercepts.

---

1 In purely formal terms, this mixture model is identical to a random effects model in which the population distribution of the parameters over households is discrete with probability mass on a small number of points in the parameter space. In fact, if the nonparametric approach of Chintagunta, Jain, and Vilcassim (1991; Vilcassim and Jain 1991) were generalized to include all model parameters, it would yield a likelihood identical to that of the Kamakura and Russell mixture model. The difference is a matter of interpretation.
If long panels were available, there would be nothing to keep us from using the method of maximum likelihood to estimate a parameter vector for each household. However, in reality, scanner panel households rarely make more than 25 purchases a year. Furthermore, each household rarely purchases all available brands during its time in the panel. These problems of insufficient data combined with household heterogeneity cause great difficulty. The MLE does not exist for many household purchase histories. Specifically, if households are observed choosing only a subset of brands, some of the choice model parameters cannot be identified. In the ERIM tuna data we examined, only two households of 777 were observed to have purchased all brands in the choice set at some point in their panel history. For a household with an incomplete purchase history, the maximum likelihood procedure will set some of the parameters to infinity to drive the fitted probabilities to zero for the brands that have not been purchased. We do not set the parameters at infinity in practice, of course, but merely shut down the routine that maximizes the likelihood after a specified tolerance level is reached. This method results in what are called extended maximum likelihood estimates (algorithm 3 of Clarkson and Jennrich 1991). In a sense, these extended maximum likelihood estimates are inconsistent with the basic choice model as applied to sets of brands of the same product. We believe it is fundamentally incorrect to infer on the basis of only a few observations that the household will never purchase certain brands regardless of the prices of the other brands.

Though we would not infer that a household has a limited choice set on the basis of only a few observations, we do not rule out the conceptual possibility of limited choice or “consideration sets.” According to the consideration set concept (see Hauser and Wernerfelt 1989; Roberts and Lattin 1991), some households have a limited choice set outside of which they will never purchase regardless of the price of outside alternatives. For example, some households may never purchase tuna packed in oil even if it were available at zero price. Typically, consideration sets are constructed from past purchase behavior and then used to specify a reduced set of brands for a logit model. Our panel dataset provides evidence that this practice is not appropriate. If we develop a consideration set based on the first two thirds of each household’s purchase history, 17% of the households purchase outside the consideration set in the final third of the data. This finding is not caused by short panel histories. If we limit attention to households with 15 or more purchases, more than 20% purchase outside a consideration set. Fundamentally, with only 10 or so purchase observations, it is impossible to say whether a household has zero choice probability for an item or merely has small probability of choice.

Bayesian Approaches to Fixed Effects Estimation

All current fixed and random effects approaches to modeling heterogeneity with short panels rely on some sort of a priori restrictions on the extent of variation in the parameters across households. We propose an alternative Bayesian method that introduces additional information through a prior distribution. We use a prior constructed from a pooled sample in which all households are assumed to have the same choice model parameters. The data for each household move the parameter estimates away from the pooled estimates in a natural Bayesian updating process.

As an illustration of our approach, consider the Bayesian analysis of the data from household i facing m choice alternatives. Let \( l(\theta | \text{Data}_i) \) denote the likelihood for household i and \( \theta \) denote the model parameters. For the standard logit model, \( l(\theta | \text{Data}_i) = \Pi_{t=1}^{T_i} p_{it}^{-z_t} \cdot \prod_{j=1}^{m_i} p_{ij}^{-z_t} \) where \( p_{it} = \exp [x_{it}' \theta] / \Sigma_{j=1}^{m_i} \exp [x_{ij}' \theta] \), \( z_t \) denotes the vector of multinomial outcomes for household i on purchase occasion t, and \( x_{ij} \) corresponds to the covariates for household i, alternative j, and purchase occasion t. We postulate a prior distribution over \( \theta \), denoted \( \pi \). This prior is combined with the likelihood to form the posterior for household i.

\[
\pi(\theta | \text{Data}_i) \propto l(\theta | \text{Data}_i) \pi(\theta)
\]

It is common for Bayes estimators to be constructed by using the mean of the posterior distribution. The sampling properties (as measured by bias, variance, mean squared error) of the Bayes estimator can be analyzed and compared with those of other proposed estimators such as the MLE.

Any Bayesian analysis depends on a convenient and reasonable representation of the prior information. In our case, the prior specification is doubly important because the sample size is small and hence the prior will influence the estimator. Indeed, as discussed subsequently, the prior is important to achieve parameter identification on the household level when the household is observed to choose from a subset of the available choices. On an intuitive level, we propose that prior information be used to ensure that households always have a positive, but possibly small, probability of choosing any alternative. Researchers who endorse a consideration set approach will find this assumption to be a limitation of our analysis. The prior information is coupled with household sample information to form the household estimate.

A Formal Bayesian Approach

Bayesian approaches to parameter estimation have been criticized on two grounds. Critics argue that the prior

\[\text{footnote text}\]
distribution is often difficult to specify and Bayes estimators can be difficult to compute. We propose simple methods of avoiding these problems.

We propose a general method of constructing prior distributions that does not dominate the information contained in the household likelihood. The resulting default or "reference" prior is designed for situations in which the investigator is reluctant to explicitly specify a prior distribution. This default procedure can be used in nearly any situation, including the analysis of new choice models for which likely ranges of parameters may not yet be known.

The prior distribution is assumed to be multivariate normal with mean \( \hat{\theta} \) and covariance \( \Sigma \) based on pooled estimates of model parameters. The normal form of the prior can be justified as an approximate posterior distribution from some previous data analysis. For samples of moderate and large size, Zellner and Rossi (1984) show that a normal approximation to the logit posterior distribution is highly accurate. Because we use the entire panel sample to assess the prior, we are justified in using the normal approximation. It is only when the individual-household likelihood function is introduced that we must worry about the accuracy of normal approximations.

In the spirit of an empirical Bayes approach, information across households is used to estimate or elicit parameters of the prior. All households are pooled into one large sample under the assumption that all parameters are equal across households. The pooled likelihood is then maximized to yield the pooled MLE, \( \hat{\theta} \), and the estimated variance-covariance matrix \( \Phi \). The prior mean is set equal to the MLE (\( \theta = \hat{\theta} \)), whereas the prior covariance matrix is set proportional to \( \Phi \).

At this point, we have reduced the problem of specifying a complete prior distribution to the choice of the scaling of the prior covariance matrix. We interpret the scaling parameter of the prior as the size of an imaginary sample (see McCulloch and Rossi 1991 for a detailed justification of this idea). In general, our prior with a sample size parameter of \( \nu \) is \( \text{MVN}(\theta = \hat{\theta}, \Sigma = N/\nu \Phi) \). For example, the prior covariance matrix can be scaled to represent a sample of size one by setting \( \nu \) to one.\(^3\) A prior set with a sample size parameter of one imparts only a modest amount of prior information to keep the parameter estimates from drifting off to infinity. For a prior with a sample size of one, the household-level estimates will be determined largely by the data for that household only. Hence households with only a small number of purchase occasions will have very imprecise parameter estimates. Setting the prior sample size to larger numbers will produce estimates that are "shrunk" toward the pooled estimate, especially for households with a very small number of observations. As the prior sample size is increased beyond the number of observations for a household, the household estimates will converge to the homogeneous or pooled model estimates. In the application section we investigate the prior sensitivity and provide estimates of the posterior standard deviation for various priors and various numbers of purchase observations.

Some researchers may regard the sensitivity of the household parameter estimates to prior tightness as a weakness of our approach. We believe the flexibility and interpretability of our prior are strong points. Very small prior sample sizes will put little weight on the information contained in the purchase records of other households in formulating a household estimate. Large prior sample sizes result in household estimates that put more weight on the information from other households and, thus, assume more commonality or homogeneity in the population of households. In many applications it will be useful to specify truly informative priors that are fairly tight. There are no automatic procedures for choosing the prior sample size parameter that will perform well in all situations. The user must carefully consider the prior and perform prior sensitivity analysis.

**Calculating the Formal Approach Estimates**

Household parameter estimates are obtained by combining the prior with the household likelihood to form the household posterior as indicated in equation 1. One popular Bayes estimator is the mean of the posterior distribution. However, there is no analytic expression for the posterior mean. These posteriors are based on multinomial likelihoods that are nonlinear in the parameters. We approximate the posterior mean by the posterior mode.\(^4\)

The posterior mode represents a compromise between the sample information represented in the household likelihood and the prior information. The MLE, if it ex-

\(^3\)The amount of Fisher or expected information in a sample of size one can be computed easily from an estimate of the information in a sample of size N. We recognize that the information matrix for a random sample of size N is simply N \( \Phi(\theta) \) where \( \Phi(\theta) \) is the average information contained in one observation. We deflate the estimate of the information matrix of the pooled sample so that it is an estimate of the expected information in one observation by multiplying each element of the matrix by 1/N.

\(^4\)The posterior mean is commonly used as an estimator. We use the posterior mode as an approximation to the posterior mean. A more accurate estimate might be constructed by using approximations proposed by Tierney and Kadane (1986) instead. Poirier and Koop (1991) show that though the Tierney and Kadane methods will enhance accuracy, there is only a minor improvement over the posterior mode. The posterior mode is the maximum of the function formed by summing the log likelihood and the log prior. It is important to note that the log of the normal prior is strictly concave and has a unique maximum. In addition, the log likelihood is bounded above by zero. In the case of the standard logit model, the log likelihood is also strictly concave. Thus, the posterior mode is guaranteed to exist and to be unique. The same is not true of the log likelihood itself, which may not have a unique maximum.
ricks, will be "shrunk" toward the prior mean vector. It is important to note that the posterior mode will always exist even when the household MLE does not. The role of the normal prior is twofold, (1) to provide a centering point toward which the sample information is shrunk and (2) to ensure the existence of a posterior mode. In practice, we use a relatively diffuse prior so that the sample information for each household will greatly influence the posterior mean. The damped tails of the normal prior will achieve identification even in the cases in which the likelihood has no unique maximum. The normal prior keeps the parameters from drifting off to infinity and keeps the associated fitted probabilities from approaching too close to zero or one.

Standard maximization routines can be used to calculate the posterior mode for each household. The code used to find the maximum of a logit likelihood function can be modified with little effort to produce our Bayesian household-level estimates. To see this, recall that from equation 1 the log of the posterior is simply the sum of the log likelihood and log prior, \( \log(\pi(\theta|\text{Data}) = \log(l) + \log(p(\theta)) \). Thus, the objective function, gradient, and hessian from the standard likelihood analysis need to be modified only by adding a second term that comes from the prior. Equations 2, 3, and 4 give the exact formulas.

\[
(2) \quad \log(\pi(\theta|\text{Data})) = L(\theta) - 1/2(\theta - \bar{\theta})'\Sigma^{-1}(\theta - \bar{\theta}) + K
\]

\[
(3) \quad g_n = g_L - \Sigma^{-1}(\theta - \bar{\theta})
\]

\[
(4) \quad H_n = H_L - \Sigma^{-1}
\]

\( \theta \) is the prior mean, which we take to be the MLE based on the pooled sample of all households, and \( \Sigma \) is the prior variance-covariance matrix; \( g_n, g_L \) are the gradients of the posterior and log likelihood, respectively; and \( H_n \) and \( H_L \) are the hessians of the posterior and log likelihood. Finally, \( K \) is the log of the integrating constant of the posterior distribution. The \( i \) subscripts have been dropped for simplicity. Every standard optimizer (such as NPSOL or GQOPT) requires that the user supply objective function, gradient, and (sometimes) hessian routines. It is a simple matter to add code from the preceding equations. Self-contained FORTRAN code and UNIX compilation scripts are available on request from the authors.

Having laid out the theory and computation methods for our Bayesian approach, we now apply the approach to both actual and simulated panel data. We compare the Bayesian approach with present methods of adjusting for heterogeneity.

\[\text{For the small sample sizes encountered in scanner panel data, standard estimates of the hessian such as the outer product approximation used in the BHHH quasi-newton method can be very inaccurate. For the standard logit model, the exact hessian is simple to compute and we have used the exact hessian in deriving the posterior modes used throughout the article. We also made comparisons with the outer product method and verified the inaccuracy of the outer product approximation.}\]
APPLICATION TO SCANNER PANEL DATA

We explored the Bayesian estimator with actual scanner panel data. Our goals were both substantive and methodological. On the substantive side, we investigated the extent of household heterogeneity detected in the data. In particular, we emphasized the importance of detection and measurement of heterogeneity in nonintercept parameters such as price, display, and feature response. Methodologically, we assessed the predictive performance of our estimators and compared them with traditional loyalty-based adjustments advocated previously (Allenby and Rossi 1991b; Guadagni and Little 1983).

The scanner panel dataset consists of household tuna purchases in three grocery chains in Springfield, Missouri. Seven tuna UPCs from four brands, which account for more than 87% of the total category volume, are included in the dataset. The brand names are Chicken of the Sea, Starkist, 3 Diamond, and a house brand. The nationally distributed brands include both water- and oil-packed tuna, but only the water-packed house brand was included for analysis. The tuna panel data serve well to test the performance of our techniques with very short panels. The median number of purchase occasions in our data is eight per household with 25% of the households reporting six or fewer purchase occasions and 75% reporting 11 or fewer purchase occasions. There are 777 households, observed over a maximum of 2.5 years, in the dataset. In addition to information on competitive pricing, display and feature information is captured in the dataset.

Comparison With the MLE and Prior Sensitivity

Before we discuss our substantive findings for ERIM panel purchases of tuna, we use the panel data to shed further light on the performance of our procedure in relation to estimating logits for each household. In addition, we illustrate the sensitivity of the household estimates to the choice of the prior tightness parameter (the sample size parameter, \(v\)). Table 1 gives household estimates of the price coefficient for 15 randomly selected

<table>
<thead>
<tr>
<th>Household</th>
<th>No. of observations</th>
<th>Individual logit (^a)</th>
<th>Bayesian approach with reference prior sample size of (v)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>-40.84</td>
<td>-8.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.41)</td>
<td>(3.83)</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>-45.27</td>
<td>-9.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.26)</td>
<td>(5.52)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.85)</td>
<td>(2.99)</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>-116.7</td>
<td>-9.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.11)</td>
<td>(2.05)</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>-25.22</td>
<td>-10.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.15)</td>
<td>(6.01)</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>-8.21</td>
<td>-6.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(98)</td>
<td>(98)</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>-21.16</td>
<td>-7.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.64)</td>
<td>(2.56)</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>-44.65</td>
<td>-6.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.32)</td>
<td>(2.09)</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>-6.51</td>
<td>-11.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.47)</td>
<td>(4.41)</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>-5.27</td>
<td>-11.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.29)</td>
<td>(1.28)</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>-4.91</td>
<td>-6.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.91)</td>
<td>(3.51)</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>-25.19</td>
<td>-6.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.47)</td>
<td>(4.18)</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>-14.85</td>
<td>-10.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.36)</td>
<td>(3.73)</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>-5.78</td>
<td>-4.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.83)</td>
<td>(2.72)</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>-52.65</td>
<td>-21.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.24)</td>
<td>(5.28)</td>
</tr>
</tbody>
</table>

\(^a\)Posterior standard deviation is in parentheses.

\(^b\)Logit model includes display, feature, and brand intercepts for a total of nine parameters. It is estimated using convergence tolerance = 1.0 \(\times 10^{-6}\); note that the standard errors for the individual logits are not defined because of a singular hessian (incomplete histories).

\(^c\)Price coefficient prior standard deviations = 11.01, 7.79, 4.92, 3.48, 1.10.
households. The second column lists the number of purchase occasions for each household. As emphasized previously, many households had only a small number of occasions whereas some households had a large number. The third column gives the estimates of individual logit models using the extended MLE procedure outlined previously. The fourth through seventh columns report the estimates based on priors of varying degrees of tightness, from a one-observation prior to a 100-observation prior. Below the Bayes estimates are approximate posterior standard deviations, calculated from the inverse of the posterior hessian.

The table clearly shows the problem with the extended MLE. Recall that for households with incomplete choice histories, the standard MLE is undefined for all model parameters (not just the intercepts). All households but two in our panel have incomplete purchase histories based on our choice set of seven. The extended MLE procedure initiates the search for a maximum and terminates after the percentage change in the likelihood is below a specified tolerance. The extended MLEs of the intercepts and price coefficient drift off to infinity. Many of the point estimates of the price coefficient are absurdly large (with a smaller tolerance, we could drive them all beyond any specified number). In addition, the identification problem renders the hessian of the likelihood singular and there are no asymptotic standard errors to report for the extended MLE.

Table 1 also illustrates the sensitivity of the Bayes estimates to the prior. For a prior sample size of 100, all estimates are close to the prior mean (the Guadagni and Little estimate of \(-6.0\)), whereas the estimates for a prior sample size of one are very variable. For small prior sample sizes and households with a small number of purchase occasions, some posterior standard deviations are large. This finding merely indicates that if there is little information in the sample for the household and the investigator is not willing to impart strong prior information, estimating household parameters will be difficult. In our view, the one-observation prior is overly diffuse for many applications. The prior standard deviation is \(11.01\), which means that the prior mass is spread over the interval \(-39\) to 27. Most analysts would agree that the price coefficient is unlikely to be positive and that values of \(-20\) or smaller would imply overly sensitive switching behavior. A more reasonable prior, which is still diffuse but rules out these unrealistic values, corre-

responds to the reference prior size of five. In the analysis reported next, we use the five-observation prior on the grounds that it more reasonably represents our information. In addition, as we show, Bayes estimates based on the five-observation prior outpredict estimates based on one- and two-observation priors.

**Measured Parameter Heterogeneity**

The household estimates computed by our formal Bayesian approach reflect the considerable amount of household heterogeneity in all model parameters. Our logit model is a standard one-price logit (using log price and display and feature dummy variables equal to one if the item is displayed anywhere in the store or featured in a major ad in the local newspaper). Summary statistics on the distribution of logit model log price, display, and feature coefficients across households follow.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Mean</th>
<th>S.D.</th>
<th>Sampling error component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>(-6.85)</td>
<td>1.78</td>
<td>0.54 (30%)</td>
</tr>
<tr>
<td>Display</td>
<td>(0.34)</td>
<td>0.71</td>
<td>0.30 (42%)</td>
</tr>
<tr>
<td>Feature</td>
<td>(0.57)</td>
<td>0.67</td>
<td>0.14 (21%)</td>
</tr>
</tbody>
</table>

The mean of the price slope distribution is \(-6.85\) with a large standard deviation of 1.78. The feature slope coefficient shows even more heterogeneity with a larger standard deviation in relation to the mean.

A legitimate criticism of these standard deviation estimates is that sampling error may cause them to be inflated. To see this point, suppose there were no heterogeneity at all in slope parameters. Estimation error will induce variation in the distribution of estimates unrelated to the true heterogeneity. To gauge the seriousness of this problem, we estimate the proportion of parameter variation attributable to sampling error. We use the asymptotic covariance matrix from the pooled model and inflate the sample variance to account for the much smaller average sample size for each individual household. The right column in the preceding table gives these estimates of the standard deviation in the absence of heterogeneity. In parentheses are the estimated percentages of the total variation that can be ascribed to sampling error. Clearly sampling error alone cannot account for the observed variation in slope coefficients. Less than half of the variation is due to sampling error. In addition, the proportion of total variation explained by sampling error is lowest for the feature variable, suggesting greater heterogeneity in household response for features than in that for displays and promotions. This finding seems reasonable in light of the varied subscription and reading habits of households in the population. Also note that these estimates of the proportion of variation due to sampling error are overstated, because the Bayesian estimator certainly has better sampling performance than the MLE.

It is logical to ask whether the heterogeneity in the response coefficients can be related to measured demographic variables. If so, interaction terms could be added to the models to capture this heterogeneity without re-

---

*The posterior standard deviations computed in Table 1 are based on standard asymptotic approximations. To check this asymptotic approximation, we computed estimates of the exact finite sample posterior standard deviations using numerical integration. We used an importance sampling method that has become standard in the Bayesian econometrics and statistics literature (see Zellner and Rossi 1984). Though there are some differences between the exact and approximate standard deviations, there is no systematic pattern in which the exact finite sample results tend to be larger or smaller than the asymptotic approximation.*
sorting to more complicated fixed effects or random effects models. We regressed our estimated price coefficients against the log of household income, log of family size, and dummy variables for household head college education and retirement status. This fairly standard set of variables accounts for less than 2% of the total variation in the price coefficients across households. This finding is consistent with those of previous research that show loyalty measures dramatically improve the fit of choice models even after demographic variables are entered.

Figure 2 is a histogram of the price coefficients estimated for the one-price logit model. There is considerable household heterogeneity in price sensitivity. Furthermore, the distribution is clearly unimodal and reasonably symmetric, suggesting that there is no evidence of distinct segments of price sensitive and insensitive households.

Implications for Couponing Behavior

In tuna and other product categories, both store and manufacturer coupons are used as a method of stimulating sales. Coupons may provide a means to price discriminate between price sensitive and price insensitive consumers as advanced by Narasimhan (1984). Because our procedures generate household-level estimates of price sensitivity, a first step toward a better understanding of couponing behavior would be to relate price sensitivity estimates to coupon usage by the households. Using the logit model’s price coefficient, we sorted the households in ascending order from the most price sensitive to the least. We then looked at whether or not the households used coupons during the entire panel period. Thirty-eight percent of the households in the first and highest quartile of the price coefficient distribution used at least one coupon whereas only 20% of those in the lowest quartile used coupons. This finding strengthens the view that couponing can be used to attract price sensitive consumers. In addition, it provides independent corroboration of the quality of our price coefficient estimates; note that coupon usage data were not used to compute the price coefficients.

One potentially valuable use for our estimated model coefficients would be to target a coupon distribution at specific households. For example, consider the ideal drop of coupons for the Chicken of the Sea brand. We expect a strong response from households that show loyalty toward other brands and yet are price sensitive. Ideally, we would like to distribute coupons to such households. Within the standard logit parameterization, we can implement this drop by targeting households for which the intercept of some other brand exceeds the intercept of Chicken of the Sea oil or water items and whose price coefficients rank above the third quartile level. Of the 777 households, 109 fit into this category of households potentially most responsive to a Chicken of the Sea coupon drop. To estimate the potential gains from targeting, we compare the response to a targeted drop to the 109 susceptible households with that from a blanket coverage of all households in our sample. For illustrative purposes, we consider a 30-cent coupon for Chicken of the Sea tuna in water. Note that the identification of a target group of households could be approached by estimating a finite mixture model, identifying a “coupon-susceptible” segment, and then using a classification rule to assign households to this segment. This procedure is fundamentally different from the one advocated here. An interesting avenue for future research would be to examine the relative efficiencies of these procedures for generating the target population.

To estimate the effects on eventual sales from a coupon drop is a difficult task. Some fraction of the households will not be aware of the coupon. In addition, some cost of exercising the coupon must be factored into the net effect of the coupon. For the purposes of our calculations, we assumed that for households that are aware of the coupon and willing to use it, the coupon acts as a temporary price cut in the full amount of the coupon. To gauge the difference in the response of the two sets of households to the coupon, we computed the average predicted probability of purchase of Chicken of the Sea tuna in water at the modal price of 89 cents and compared it with the predicted probability of purchase at the reduced price net of the coupon. For the population of all households, the median predicted probability increased from .35 to .67 after the coupon drop. For the targeted population, the predicted probability of purchase increased from .20 to .71. Thus, under the targeted coupon distribution, we will persuade many households that otherwise might not purchase Chicken of the Sea to switch from other brands. A major advantage of the targeting over blanket coverage is that price discounts will not be given away to households that are
very likely to purchase Chicken of the Sea without the discount.

The ultimate profitability of these sorts of targeted couponing strategies depends on many factors (see Blattberg and Neslin 1990, chapter 10, for an excellent overview of couponing costs). With current direct mail technology, the cost of reaching the targeted population is probably substantially greater than that of reaching the entire population with a newspaper insert or blanket mailing. However, the costs of targeting will certainly decline in the future with the advent of point-of-purchase couponing and improved technology. For high ticket items, the benefits of targeting may far exceed the cost differential. The key to the implementation of these strategies is the availability of historical data on household purchases coupled with techniques for producing household-level parameters.

Implications of Heterogeneity for Aggregate Market Response

The distribution of estimated parameters across households is interesting, but tracing the implications of this heterogeneity for managerial decision making is not straightforward. In this section, we examine the implications of heterogeneity for the computation of aggregate market share responses to changes in price, display, or feature. Expected aggregate market share is simply the sum over all households of the purchase probabilities divided by the number of households. Because there is a large number of households, expected and actual market shares will be very close by the law of large numbers. Aggregate chain or store-level pricing decisions are made, in part, on the basis of estimates of the derivative of market share with respect to various marketing mix variables. Manufacturers want to know how their market share will respond to price promotions, displays, or features. The derivative of market share is the sum of the derivatives of the probabilities of purchase for each household.

\[
\frac{\partial \text{Mkt Share}(k)}{\partial x_i} = \sum_{j=1}^{N} \frac{\partial}{\partial x_i} \Pr(k|x = x_i, \theta)/N^* 
\]

Here \( x \) is the vector of marketing mix variables, log price, display, and feature. \( N^* \) is the total number of observations in the panel. In equation 5, we sum over both households and time, thus averaging the response over the distribution of prices faced by the households and eliminating the need to pick a point at which to compute the elasticity.

Estimates of aggregate market share derivatives are influenced dramatically by the degree and type of heterogeneity allowed in the \( \theta \) vector. Table 2 gives estimated derivatives for various estimation procedures in the one-price logit model. In the first column are derivatives based on a pooled \( \theta \) estimate that is common across all households (the case of no heterogeneity). The second column allows for intercept heterogeneity by use of the Allenby/Rossi procedure for constructing a loyalty measure. The price elasticities show an approximately 10% reduction whereas the display and feature derivatives show little change. The role of heterogeneity can be better understood from the expression for the derivative in the case of the standard logit model: \( \frac{\partial \Pr}{\partial \log \text{price}} = \beta_0 \Pr(1 - \Pr) \). \( \Pr \) denotes the choice probability locus viewed as a function of price. Adjusting for intercept heterogeneity increases the common price coefficient estimate from -5.38 to -5.95. Adjusting for both slope and intercept increases the average price response further to -6.85. This procedure would tend to increase the size of the price elasticity through the first term. In contrast, allowing the intercepts to vary from household to household and increasing responsiveness to price has the effect of making the fitted probabilities more flexible and the probabilities move toward zero or one, which tends to lower the magnitude of the elasticity. In the case of the price elasticity, it is the latter effect that dominates.

The third column of Table 2 are market share derivatives for the Bayes household estimates that allow for heterogeneity in both the intercept and the slope with respect to price, feature, and display. Allowing for slope heterogeneity has dramatic effects on the size of the derivatives, particularly for the feature variables. The estimated response to features nearly doubles when slope
heterogeneity is allowed. An investigator who fails to take into account heterogeneity will seriously undervalue feature advertising in this dataset. The fact that the feature derivative estimates are most influenced by heterogeneity is consistent with our previous finding of more heterogeneity in the feature slopes than in the other model slopes.

**Predictive Performance**

We now compare the predictive performance of procedures that allow intercept heterogeneity only through the use of loyalty measures with our Bayesian methods that accommodate heterogeneity in all parameters. Predictive performance is not the only or the ultimate test of a model or estimation procedure. As is well known, particular models may forecast well simply because of a reduction in the number of parameters even though the predictions may be biased. Perhaps even more important is the realization that superior predictive performance in data similar to the data used for prediction does not mean that the model will predict well after a change in policy (such as a change in the pricing policy). Further, very simple models that predict well may not yield sufficiently rich detail to explain or better understand individual household behavior. Our goal in examining predictive performance is to ensure that we have not paid too high a price in predictive performance for the very rich detail of household behavior available through our method.

Measuring predictive performance in the choice setting is difficult and controversial. To provide a balanced view of performance, we report three standard measures and introduce a new measure. In addition to predictive log likelihood, hit rates, and residual mean absolute deviation, we report the average probability for "hits."

The predictive log likelihood is a very conservative measure of performance. It assesses a high penalty for low predicted probabilities in cases when the household is observed to make purchases. These highly penalized "misses" tend to overshadow improvements in the probability of "hits" due to the Bayesian approach. One can argue that from the point of view of policy evaluation, a few misses on the household level are not nearly as important as more accurate prediction of the bulk of household response to new pricing or advertising policies. For this reason, many researchers advocate the use of a hit rate—the percentage of time that observed choices have the highest fitted probability. Hit rate is a more balanced measure that does not penalize incorrect predictions as heavily as predictive log likelihood. A defect of the hit rate measure is that it does not provide an assessment of the confidence with which the correct predictions are made. What is most desirable is a procedure that produces very high fitted probabilities for the choices made. To assess the predicted probabilities for the alternatives chosen, we introduce the average probability for hits as a supplement to the hit rate. Finally, we compute the average absolute value of the "residuals" (MAD), where the residual is defined as the choice vector (vector of ones and zeros) minus the predicted probabilities.

Table 3 reports the results of predictive tests. The dataset is broken into thirds, with the final third reserved for predictive testing. The first third of the data is used to fit a model without a loyalty parameter. Two different varieties of loyalty measures are employed to allow for intercept heterogeneity. We consider both the Guadagni and Little (1983) and the Allenby and Rossi (1991b) approaches. To further bolster our choice of the five-observation prior, we consider Bayes estimates based on the one-, two-, five-, 10-, and 100-observation priors. The second third of the data is then used to reestimate the model parameters with a loyalty parameter. As might be expected, the in-sample log-likelihood values for our household-level Bayes estimates are approximately half of the log-likelihood values for the Guadagni and Little loyalty approach. The key question is whether the introduction of a large number of household parameters causes the predictive performance to deteriorate.

The five-observation-prior Bayes estimates outperform the loyalty-based procedures on all four performance criteria. In general, the Bayes estimates produced by the five-observation prior have the best predictive performance. Even with a strong 100-observation prior, the Bayes estimates outperform standard loyalty adjustments. The reason is that even the 100-observation prior accommodates some slope heterogeneity not allowed for in the standard procedures. For example, the price coefficient estimate from the 100-observation prior ranges from -5.0 to -7.0.

The superior predictive performance of the Bayes estimates is achieved in spite of the introduction of more than 1500 separate household slope parameters. In general, there is a tradeoff in predictive performance between the increased flexibility imparted by highly pa-

---

```
<table>
<thead>
<tr>
<th>Heterogeneity adjustment procedure</th>
<th>Log likelihood</th>
<th>Hit rate</th>
<th>Average hit probability</th>
<th>Residual MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guadagni/Little</td>
<td>-2574</td>
<td>72.1</td>
<td>.580</td>
<td>.081</td>
</tr>
<tr>
<td>Allenby/Rossi</td>
<td>-2647</td>
<td>70.3</td>
<td>.582</td>
<td>.084</td>
</tr>
<tr>
<td>Bayes estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 observation</td>
<td>-2693</td>
<td>70.3</td>
<td>.624</td>
<td>.078</td>
</tr>
<tr>
<td>2 observation</td>
<td>-2640</td>
<td>72.0</td>
<td>.652</td>
<td>.071</td>
</tr>
<tr>
<td>5 observation</td>
<td>-2513</td>
<td>72.5</td>
<td>.641</td>
<td>.072</td>
</tr>
<tr>
<td>10 observation</td>
<td>-2433</td>
<td>71.9</td>
<td>.608</td>
<td>.078</td>
</tr>
<tr>
<td>100 observation</td>
<td>-2506</td>
<td>71.9</td>
<td>.593</td>
<td>.081</td>
</tr>
</tbody>
</table>
```

---

\(^3\)Counting the number of household parameters in a Bayesian approach is somewhat misleading. The use of informative priors pulls the estimates toward a common mean. With very informative priors, the effective number of parameters is undoubtedly much less than 1500. In the limiting case of an extremely tight prior, the number of parameters is the same as the number of elements in the pooled parameter vector.
rameterized models and the sampling error introduced by a larger number of estimated parameters. However, in this case, the incorporation of slope heterogeneity via the Bayes estimates captures an important aspect of the data and is balanced by the sampling error imparted by the larger number household parameters. The predictive performance of the Bayes estimates corroborates our findings in Table 1 that a substantial amount of heterogeneity in the slope coefficients is present in the data.

CONCLUDING REMARKS

For scanner panel data, the accurate estimation of household parameters in models of consumer behavior is difficult. The number of observations for an individual household is often very small (<25) and households rarely exhibit enough choice variation to allow for the estimation of all parameters. For example, the fact that households seldom choose all brands in the choice set results in unidentified intercepts for the brands not selected.

Additional information is needed to identify the parameters that describe household behavior. We present a Bayesian method of introducing additional information into the estimation procedure through the prior distribution. A reference prior is proposed that aids in the implementation of the procedure. The effect of our procedure is to smooth fitted probabilities away from zero or one.

Our procedure has three basic advantages: (1) household estimates for all parameters, including slope parameters, are produced, (2) the methods can be applied by making minor modifications in present optimization routines, and (3) our methods can be applied to any model, not just the multinomial logit model considered here (see Allenby and Rossi 1992 for applications to other choice models and a discussion of an alternative Bayesian approach).

In contrast to parametric random effects models, the estimators do not require a priori specification of the distribution of heterogeneity. Neither the family of the distribution (e.g., normal, gamma, etc.) nor its relation to observed household characteristics needs to be identified. Hence, the estimator is particularly well suited for situations in which little is known about the determinants of demand and investigators are interested in approaching the data inductively.

We apply our procedures to panel data on tuna purchases and find strong evidence of both slope and intercept heterogeneity in the dataset. The degree of heterogeneity is three to 10 times the level expected from sampling error in the household estimates. By some measures, there is considerably more heterogeneity in the feature response coefficients than in the price or display response coefficients. Moreover, the slope heterogeneity dramatically changes aggregate estimates of market share response to marketing mix variables. Procedures that use only loyalty measures to account for heterogeneity will produce very biased estimates of market share response. By properly accounting for slope heterogeneity, we find that the response to feature advertising is much larger than what would be predicted from an intercept-only approach.

We also illustrate the value of household estimates by showing how they can be used to devise a targeted coupon drop. This example is meant to illustrate only a few of the benefits of household parameters. We hope to pursue several directions in future research. First, the availability of household-level estimates opens the possibility of customizing the face value of the coupon to the specific household. In fact, some retailers are experimenting with point-of-purchase couponing whereby households receive coupons based on their most recent set of purchases. Second, the problem of optimal couponing can be posed in a decision theoretic setting in which the decision rules that target individual households arise from the Bayesian solution of a profit maximization problem. This approach is a subject of our current research.

In summary, the proposed Bayesian estimators can use the wealth of information contained in scanner panel data to reveal differences in individual household demand functions that are fundamental to the theory and practice of marketing. The household-level estimates can provide the basis for better targeted and more effective marketing strategies and tactics.

REFERENCES

Kamakura, Wagner A. and Gary J. Russell (1989), "A Prob-


