JAIHAK CHUNG and VITHALA R. RAO*

The authors present a general consumer preference model for experience products that overcomes the limitations of consumer choice models, especially when it is not easy to consider some qualitative attributes of a product or when there are too many attributes relative to the available amount of preference data, by capturing the effects of unobserved product attributes with the residuals of reference consumers for the same product. They decompose the deterministic component of product utility into two parts: that accounted for by observed attributes and that due to nonobserved attributes. The authors estimate the unobserved component by relating it to the corresponding residuals of virtual experts representing homogeneous groups of people who experienced the product earlier and evaluated it. Their methodology involves identifying such virtual experts and determining the relative importance they should be given in the estimation of the target person's residuals. Using Bayesian estimation methods and Markov chain Monte Carlo simulation inference, the authors apply their approach to two types of consumer preference data: (1) online consumer ratings (stated preferences) data for Internet recommendation services and (2) offline consumer viewership (revealed preferences) data for movies. The results empirically show that this new approach outperforms several alternative collaborative filtering and attribute-based preference models with both in- and out-of-sample fits. The model is applicable to both Internet recommendation services and consumer choice studies.

Keywords: consumer preference model, online recommendation service, collaborative filtering, experience products, Bayesian latent residuals

A General Consumer Preference Model for Experience Products: Application to Internet Recommendation Services

*Jaihak Chung is Professor of Marketing, School of Business Administration, Sogang University, Seoul, Korea, and Affiliate Professor, IESEG School of Management, France (e-mail: jaihak@sogang.ac.kr). Vithala R. Rao is Deane W. Malott Professor of Management and Marketing and Quantitative Methods, S.C. Johnson Graduate School of Management, Cornell University (e-mail: vrr2@cornell.edu). The authors acknowledge Compaq Systems Research Center for providing data for their research. They thank participants at the seminars at the Peking University, the Wharton School, and the Hebrew University of Jerusalem for their comments and Srinagesh Gavirneni for his suggestions for improving this article. This work was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2010-330-B00110). Fred Feinberg served as associate editor for this article.

Assume that we use a consumer preference model to predict preferences for recommending experience products such as movies. The standard consumer preference model is limited in quantifying various qualitative characteristics of a movie such as the detailed story (episode), the appearances of main characters, and its tone and mood. As an example, consider one of the most famous scenes, taken from the movie script of *Titanic* (1997), with its main characters, Jack and Rose:

> *Jack*: "Close your eyes."
>
> She does, and he turns her to face forward, the way the ship is going. He presses her gently to the rail, standing

ISSN: 0022-2437 (print), 1547-7193 (electronic)

> right behind her. Then he takes her two hands and raises them until she is standing with her arms outstretched on each side. When he lowers his hands, her arms stay up like wings.
>
> *Jack*: "Okay. Open your eyes."
>
> Rose gasps. There is nothing in her field of vision but water. The Atlantic unrolls toward her. There is only the wind and the hiss of the water 100 feet below.
>
> *Rose*: "I'm flying!"

We believe that it is almost impossible for standard choice models to fully consider the effects of main character appearances (Jack and Rose) and the beautiful scenery of the sea in the preceding scene, which may have a large impact on consumers' preferences for that movie. Although researchers can consider the effects of qualitative attributes by categorizing and quantifying them in a simpler all-binary case with dummy variables, there still will remain qualitative information on product attributes that cannot be processed by a quantitative model for at least three reasons.

First, the categorization of such qualitative attributes may result in the loss of a significant amount of information (Stangl and Berry 2000). For example, most studies on movies use genre only as an independent variable for movie stories by using dummy variables, but such a categorization does not fully convey differences among movies in the same genre. Second, the categorization method may generate measurement errors due to the subjective nature of such categorizations. For example, it would be difficult and almost impossible to create quantitative variables to describe the overall appearances of the main characters (Jack and Rose) combining varying views (e.g., clothing, facial expressions, movements, voices) as well as the dynamic motions, colors, and sounds of the sea in that scene, which may have a significant impact on consumers' preferences for that movie. Third, the categorization method is likely to generate a large number of dummy variables that cannot be accommodated in a quantitative model due to the degrees of freedom issue. From the perspective of modeling consumer choice behavior, we can classify the attributes of a product into two categories: attributes whose information can or cannot be conveyed with a set of quantitative variables in quantitative models. In this article, we simply label these as quantifiable and nonquantifiable attributes.

Standard preference models that consider only quantifiable product attributes for preference are therefore limited in predicting consumers' preferences, especially for experience products such as entertainment services, online content, and games. The larger the contribution of nonquantifiable attributes to a product utility in a choice model, the lower will be the model's predictive power. It is noteworthy that although the provision of recommendation services is important in marketing, most prediction models for recommendation services have been developed by engineering, information systems, and information science researchers (Ariely, Lynch, and Aparcio 2004; Bell, Koren, and Volinsky 2008; Breese, Heckerman, and Kadie 1998; Koren 2009) and typically involve collaborative filtering (CF) methods. We can partially attribute this situation to the prevailing practice in marketing research of using only quantifiable attributes for predicting consumer preferences. However, CF methods are not limited by the presence of nonquantifiable attributes, because they rely on other users' preference data as a whole for prediction, in contrast to marketing research methods that rely on the quantifiable product attributes.

We use the spirit of CF and develop a general consumer preference model to overcome limitations of the inability to incorporate effects of nonquantifiable attributes. For this purpose, we decompose a product's utility in a standard preference model into two parts: one part accounted for by a set of quantifiable product attributes available for choice models and the second part not accounted for due to the presence of nonquantifiable attributes. The second part is the latent residual in the consumer's utility of a standard consumer preference model. Our model captures the latent residual of a standard consumer preference model for a target consumer by relating it to the corresponding residuals of several virtual experts representing another group of customers similar in their preferences. For this purpose, we develop a Bayesian method to identify such virtual experts by clustering and then estimate the latent residuals of the virtual experts. Our model estimates the relative importance of virtual experts' residuals according to two pieces of information: (1) experts' preference similarities (how similar a target person is to the virtual expert) and (2) experts' precision levels (the inverse of the variance of residuals within a cluster). We describe our model with reference to recommendation services, which provide a better environment to highlight our modeling approach. However, we suggest this modeling approach as a general method to improve any preference model that can be applied to consumer preference behaviors (both stated and revealed), particularly when some product attributes are not easy to quantify (e.g., for entertainment services and online content) or when there are too many product attributes (e.g., for automobiles) relative to the available amount of preference data.

To test the predictive accuracy of our approach, we apply our model to two types of consumer preference data: (1) online consumer ratings (stated preferences) for Internet recommendation services and (2) offline consumer viewership (revealed preferences) for movies. We undertake an extensive comparison of our model's predictive power with that of several major CF and attribute-based preference models, and our graphic analysis of latent residuals provides model diagnostics. Our empirical analysis shows that predictions from our approach are superior to those of previous models: The prediction hit rates of our model are 49% for stated preference data compared with 44% for the best collaborative algorithm and 43% for the best attribute-based preference model. The improvement of our model is even larger for revealed preference data (88% vs. 75%).

We organize the rest of this article into four sections: First, we provide a brief review of relevant literature on recommendation models. Second, we develop two models: Model A for the prediction of consumers' preferences and Model B for the identification of virtual experts through clustering. Third, we describe results from two applications of our model: one to recommendation systems data and the other to actual choice data. Fourth, we conclude with a discussion of the advantages and limitations of our model and some directions for further research.

LITERATURE REVIEW

The extant methodological approaches to the prediction of consumer preferences for experience products in recommendation services and standard choice studies can be classified into two categories according to types of information used for preference prediction: (1) CF models based on consumer preference similarity, developed mainly in computer science, and (2) attribute-based preference or choice models based on product attribute similarity, applied almost routinely in marketing research.

The Netflix Prize competition of the 2000s accelerated the burgeoning variety of CF modeling approaches to recommendation services (Koren 2009) in the early 1990s. These models can be divided into two subcategories according to how they incorporate consumer preference similarity: memory based and model based. Sometimes called the original approach, memory-based CF models use stated preference data (ratings) of other consumers (reference consumers) as predictors for target consumers' preferences. The most popular one is the neighborhood model (Bell and Koren 2007; Breese, Heckerman, and Kadie 1998; Koren 2009). The model-based CF approach employs a variety of general models such as matrix factorization (MF) (Bell, Koren, and Slovensky 2008; Koren 2009). These models differ from memory-based CF models in that they use model parameters to capture preference similarities through a learning model rather than directly using reference consumers' data (Koren 2009; Takács, Pilászy, and Németh 2008).

An attribute-based preference model describes the utility of a product as a weighted sum of the effects of product attributes, and the weights are estimated with consumers' preference data (stated or revealed). The model is then used to estimate the total values of products that the consumer has not used or experienced and to make recommendations to the consumer. For example, Urban, Sultan, and Qualls (2000) apply a standard logit form of consumer utility using only information on product attributes. Ansari, Essegaier, and Kohli (2000) develop an attribute-based preference model by accounting for product heterogeneity with the interaction effect of observed consumer characteristics (age and gender) and product-specific parameters and estimate it with Bayesian linear regression. More recently, Ying, Feinberg, and Wedel (2006) extend the Ansari, Essegaier, and Kohli model by considering the effect of missing responses using a two-stage ordinal probit model. In this model, the first stage describes the response choice of the movies for rating, and the second stage describes the ratings themselves in terms of movie characteristics and individual characteristics to account for heterogeneity. Whereas these two studies involve stated preference data, Bodapati (2008) proposes a model for purchase choice data (revealed preference data) consisting of a two-stage purchase process (awareness and satisfaction) to account for missing responses, called unary data, by using additional information on the firm-initiated data of responses to recommendations.

Both the CF and attribute-based approaches to the prediction of consumer preferences have advantages and limitations. CF models use holistic ratings of other consumers rather than product attributes as inputs for the prediction of target consumers' preferences. Therefore, CF models can be used for any product recommendation, including experience products, but they do not provide any insight into how a consumer evaluates a product and why he or she likes it. Furthermore, the standard attribute-based preference models that use only observed product attributes and consumer characteristics are limited in recommending experience products because they cannot consider the effects of non-quantifiable attributes on the values of the products.

DEVELOPMENT OF THE VIRTUAL EXPERT MODEL

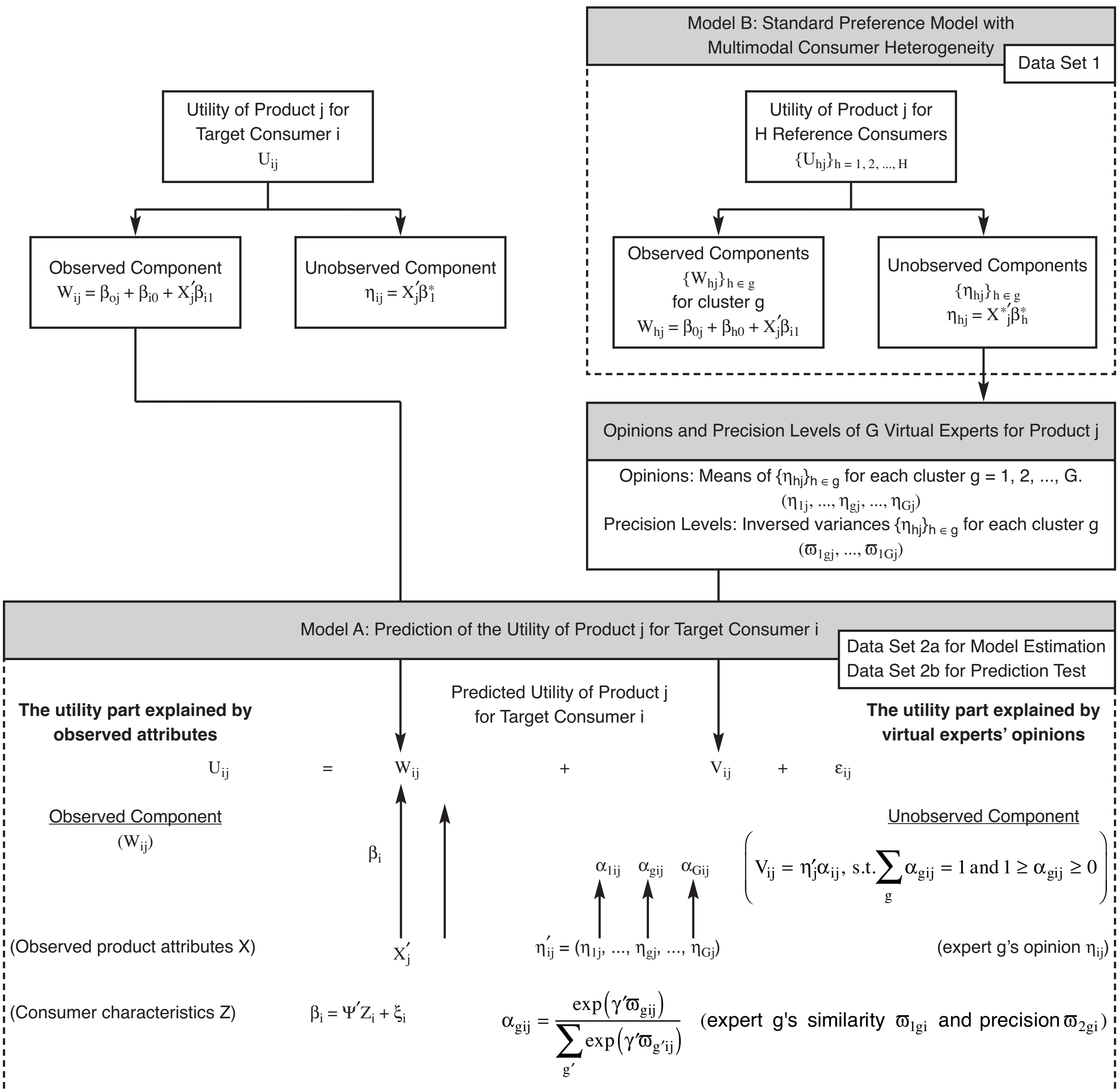
We assume that a firm offering several experience products is interested in predicting the stated or revealed preference behaviors of a group of customers, referred to as the target group, for products they have not experienced before. In addition, we assume that the firm has preference data obtained from another group of customers who have experienced some of the products, referred to as the reference group; we call this Data Set 1. We also assume that the firm has preference data for the target customers for some products; we call this Data Set 2. All people in the target and reference groups may not have experienced all the products. The sets of products the firm needs to evaluate for preference predictions may vary across people in the target group. Furthermore, we split Data Set 2 into two subsets: Data Set 2a, which is used for model estimation, and Data Set 2b, which is used for model prediction. In summary, our approach uses the preference data (Data Sets 1 and 2a) to estimate our models (Models A and B) and Data Set 2b to predict target customers' preference behaviors.

For this problem, we develop a general preference model for experience products to improve the predictive power of a standard consumer preference model, particularly when it is not easy to consider some qualitative attributes of a product or there are too many attributes relative to the available amount of preference data, by capturing the effects of unobserved product attributes with the residuals of reference consumers for the same product. We decompose the deterministic component of a product's utility into observed and unobserved components. The latter part is the residual in a standard preference model. In addition to estimating the observed component (which is routine in a choice model), we estimate the unobserved component as a linear combination of the corresponding residuals of multiple groups (clusters) of reference customers who had experienced the product previously; we call this Model A. We identify the clusters of the reference group with another model, Model B, which is a standard preference model with a finite mixture of Gaussian distributions (Allenby, Arora, and Ginter 1998; Chung and Rao 2003). Note that our general preference model uses only the commonly available types of data in recommendation and choice studies[1]: (1) consumers' stated or revealed preference data, denoted by Y; (2) product attributes, denoted by X; and (3) consumer characteristics, denoted by Z. Figure 1 provides an overview of model descriptions, their relationships, and the data sets we used for each model. We describe all details of our models in the following sections.

---

[1]We note that some previous studies in the recommendation literature have used additional data such as the date of rating (Koren 2009), firm-initiated response data (Bodapati 2008), and movie magazines' ratings (Ansari, Essegaier, and Kohli 2000).

Figure 1
THE STRUCTURE OF VIRTUAL EXPERT MODELS



Model A: The Virtual Expert Model for Preference Prediction

We let Y_{ij} represent the stated preference data, such as rating, or the revealed preference data of a target consumer i for a product j, as collected in most recommendation systems or consumer choice studies (surveys and retailer scanner systems). This measure is binary in the case of revealed preference data (e.g., buy or do not buy) or a scale with multiple points (usually ordinal) in the case of stated preference data. In general, we let R denote the number of points on this scale and r denote a specific response. We model these data by postulating the existence of (R + 1) threshold values for the latent random utilities, U_{ij}, as shown in the following:

$$Y_{ij} = r \text{ if } C_{i,r-1} < U_{ij} \leq C_{ir},$$

where r = 1, 2, ..., R; $C_{i,0} = -\infty$, $C_{i,1} = 0$; and $C_{i,R} = \infty$.We decompose the latent utility of product j for consumer i, U_{ij}, into two parts for the effects of observable and unobservable product attributes, X_j and X_j^*, respectively, as follows:

$$U_{ij} = W_{ij} + \eta_{ij}, \tag{1}$$

where W_{ij} is the observed component that can be captured using quantifiable product attributes, X_j, and η_{ij} is the unobserved component that cannot be captured by a standard attribute-based utility model. We also refer to the unobserved component as the latent residual of the standard attribute-based model.

Modeling the observed component (W_{ij}). We employ the standard preference model based on product attributes to model the observed component of the utility W_{ij}, with observed product attributes X_j, as follows:

$$W_{ij} = \beta_{i0} + \beta_{0j} + X_j'\beta_{i1}, \quad (2)$$

where X_j is a vector of observed product attributes and β_{il} is the corresponding vector of individual preference parameters. The intercepts β_{i0} and β_{0j} capture the main effects of consumer i and product j, respectively, as Ansari, Essegaier, and Kohli (2000) suggest. We set the mean of β_{0j} to zero for model identification. The last term, $X_j'\beta_{i1}$, captures the interaction effect of consumer preference and the observed characteristics of product j. Furthermore, we model the individual-specific vector of coefficients, $\beta_i = (\beta_{i0}, \beta_{i1})$, as a linear function of individual characteristics, Z_i, in a hierarchical structure:

$$\beta_i = \Psi' Z_i + \xi_{\beta i},\ \ \xi_{\beta i} \sim N(0, \Sigma_\beta)\ \forall \text{ i and j}, \quad (3)$$

where Z_i is a vector of individual characteristics including an intercept and $\xi_{\beta i}$ is the error term that accounts for unobserved heterogeneity across people. The matrix Ψ consists of the corresponding parameters and represents the effects of observed individual characteristics on individual preferences.

Modeling the unobserved component (η_{ij}). Assuming there exists a similar virtual expert g for a consumer i, we can justify our approach by showing how the unobserved component (or residual) of a product utility for consumer i is related to that for the virtual expert g. We can justify this approach using the econometricians' viewpoint by assuming the unobserved component, η_{ij}, is due to the lack of information on some unobserved product attributes denoted by X_j^* and preference parameters, β_i^*:

$$\eta_{ij} = U_{ij} - (\beta_{i0} + \beta_{0j} + X_j'\beta_i) = X_j^{*\prime}\beta_i^*. \quad (4)$$

We can conceptually relate β_i^* to the unobserved attributes preference parameter, β_g^*, for the gth virtual expert as follows:

$$\beta_i^* = \alpha_{ig}\beta_g^* + \varepsilon_i^*,$$

where α_{ig} indicates the extent of preference similarity between expert g and customer i and the error term ε_i^* can be interpreted as the degree of preference heterogeneity between customer i and expert g. Therefore, the relationship between η_{ij} and η_{gj} is as follows:

$$\eta_{ij} = X_j^{*\prime}\beta_i^* = X_j^{*\prime}\beta_g^*\alpha_{ig} + X_j^{*\prime}\varepsilon_i^* \quad (5)$$
$$= \eta_{gj}\alpha_{ig} + \varepsilon_{ij}, \text{ where } \eta_{gj} = X_j^{*\prime}\beta_g^* \text{ and } \varepsilon_{ij} = X_j^{*\prime}\varepsilon_i^*.$$

This equation indicates why we can use the residuals of other consumers to capture the residual for consumer i. We can generalize this to accommodate a set of multiple virtual experts who vary in similarity to the consumer i and estimate the unobserved component (latent residual) of a product j for target consumer i as shown in the following:

$$\eta_{ij} = \sum_g \eta_{gj}\alpha_{gij} + \varepsilon_{ij}, \quad (6)$$
$$\sum_g \alpha_{gij} = 1 \text{ and } 1 \geq \alpha_{gij} \geq 0 \text{ for all g},$$

where ε_{ij} is an error term that is normally distributed with zero mean and variance of 1. The weights α_{gij} are mixing coefficients for each of the virtual experts, and we describe how we determine them in the following subsections. This modeling structure is somewhat similar to that of a single layer mixtures-of-experts model with multiple expert networks (Jordan and Jacobs 1994); however, the experts in these models are independent submodels, whereas we determine our experts from the estimates of latent residuals of standard preference models.

Modeling mixing coefficients for multiple experts. Given that virtual experts differ in their usefulness for predicting a target consumer's preferences, we allocate weights, called the mixing coefficients, $\alpha_{ij}' = (\alpha_{1ij}, \alpha_{2ij}, ..., \alpha_{Gij})$ a to multiple experts as a logistic function of two descriptors. These descriptors are defined as follows: (1) ϖ_{1gi}, the preference similarity (membership probability) between a target consumer i and an expert g, and (2) ϖ_{2gi}, the expert precision level (standardized inverse variance of consumers' estimated latent residuals in cluster g for product j), as given in the following:

$$\alpha_{gij} = \frac{\exp\left(\gamma_g'\varpi_{gij}\right)}{\sum_{g'}\exp\left(\gamma_g'\varpi_{g'ij}\right)} \text{ for all i, g, and j}, \quad (7)$$

where $\varpi_{gij} = (1, \varpi_{1gi}, \varpi_{2gi})$ denotes an intercept, expert g's preference similarity (membership probability), and precision (inversed variance of residuals) levels. The vector $\gamma_g' = (\gamma_{g0}, \gamma_{g1}, \gamma_{g2})$ denotes the corresponding parameters for ϖ_{gij}. The intercept for the first expert (g = 1), γ_{g0}, is set to zero for model identification.

Estimating the latent residuals of reference consumers. We calculate experts' opinions, precision, and experience levels from the latent residuals of reference consumers for the corresponding products. The posterior distribution of the latent residual for reference consumer h, η_{hj}, for product j is a truncated normal density, as we show in Equation 8, in which the truncated area is conditional on preference behavior Y_{hj} and the observed component W_{hj} (Albert and Chib 1995). The latent residual drawn from this posterior is more likely to be different from the prior density N(0, 1) only when the observation is far from the predicted value determined only by the observed component.

$$\left[\eta_{hj}\middle|\beta_h, Y_{hj}, C_h\right] = \frac{\varphi\left(\eta_{hj}\right)}{\Phi\left(C_{h,Y_{hj}} - W_{hj}\right) - \Phi\left(C_{h,Y_{hj}-1} - W_{hj}\right)} \quad (8)$$
$$1\left(C_{h,Y_{hj}-1} - W_{hj} < \eta_{hj} < C_{h,Y_{hj}} - W_{hj}\right),$$

where $W_{hj} = \beta_{0j} + \beta_{h0} + X_j'\beta_{h1}$, $1(\cdot)$ is the indicator function, and $\varphi(\cdot)$ and $\Phi(\cdot)$ are the standard normal probability density function and cumulative density function, respectively.

One estimation strategy to obtain the latent residuals of reference consumers is to employ a two-stage estimation, which involves applying a standard preference model to reference customer's preference data, Y_{hj}, in Data Set 1 and then directly obtaining the residuals by subtracting Markov chain Monte Carlo (MCMC) draws for the corresponding latent utilities with those for the observed components, W_{hj}. This two-stage approach is conceptually simple and involves less computational burden in practice because reference customers' latent residuals from only the estimation of Model B, denoted by η^B_{hj}, are obtained directly from MCMC simulation for the standard preference model without an additional MCMC step. However, a drawback is that this does not take into account estimation uncertainty of experts' opinions and precision levels for the prediction of target consumers' preferences.

An alternative estimation approach is to simultaneously estimate Models A and B and reference customers' latent residuals for virtual experts' opinions and precision levels. This simultaneous estimation approach yields more efficient estimates of latent residuals for reference customers because it uses additional information on the correlation between the residuals of reference customers obtained from the application of a standard preference model, η_{hj}, and the residuals of target customers obtained from the virtual expert model (Model A), ε_{ij}. This is because these two residuals share the same missing information on the attributes, $X^{*\prime}_j$, as shown in the following:

$$(9) \qquad \eta_{hj} = X^{*\prime}_j\beta^*_h, \text{ and } \varepsilon_{ij} = X^{*\prime}_j\varepsilon^*_i$$

Therefore, the simultaneous estimation method yields more efficient estimates of latent residuals, denoted by η^A_{hj}, from the joint distribution of reference consumer h's residual, η_{hj}, and target consumers' residuals $\{\varepsilon_{ij}\}$ for product j.[2]

The simultaneous estimation approach enables us to obtain more efficient estimates of reference consumers' residuals but involves a significant computational burden, while the two-stage estimation approach provides less efficient estimates but requires much less computational burden. To demonstrate the versatility of our general preference model, we employ the simultaneous estimation approach for two empirical applications with stated and revealed preference data (for details on how to estimate reference customers' latent residuals using MCMC simulation procedures, see Web Appendix A, Steps 7 and 8, for the simultaneous estimation approach and Web Appendix B, Step 6, for the two-stage estimation approach as an option; www.marketingpower.com/jmr_webappendix).

Missing responses. Most recommendation systems use databases consisting of online user ratings for a small number of products, which users sometimes chose on the basis of their consumption experience or by recommendation systems based on their nonrandom selection process (Bodapati 2008; Ying, Feinberg, and Wedel 2006). Our model accommodates this problem of missing not completely random responses (Little and Rubin 1987) by inserting a "response choice" stage, as Ying, Feinberg, and Wedel (2006) suggest. Consumers' response behaviors can be regarded as another type of choice behavior (i.e., a consumer choosing to respond to a preference question). Let PS_{ij} denote the probability of individual i's response choice of product j. With X_j denoting a set of an intercept and covariates consisting of product attributes, and the disturbance term, ε^s_{ij}, following a normal distribution with mean zero and unit variance, we can model the corresponding latent utility of response choice U^s_{ij} as follows:

$$U^s_{ij} = X^{\prime}_j\beta^s_i + \varepsilon^s_{ij}.$$

If consumer i's response choices are related to his or her preference behaviors, preference behavior probability PR_{ijr} depends on the consumer rating product j, as we show in the following equation (Ying, Feinberg, and Wedel 2006):

$$(10) \qquad E\left(U_{ij}\middle|DS_{ij}=1\right) = E\left(W_{ij}+V_{ij}+\varepsilon_{ij}\middle|\beta^s_iX_{ij}+\varepsilon^s_{in}>0\right) = W_{ij}+V_{ij}+\rho_i\frac{\varphi\left(\beta^s_iX_j\right)}{\Phi\left(\beta^s_iX_j\right)},$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the corresponding cumulative density function and probability density function, respectively. Its inclusion is optional and depends on the way data are collected. We use the general form in our first empirical application (online recommendation data) by including the effects of such nonignorable missing responses (Little and Rubin 1987), but not for the second empirical study (survey data), because it contains no missing responses. Finally, the Appendix summarizes the virtual expert models, and Web Appendixes A and B (www.marketingpower.com/jmr_webappendix) describe the MCMC simulation for their estimation.

Model B: The Identification of Virtual Experts

We use the averages of the residuals of a group of reference customers (clusters) with relatively similar preferences to capture the unobserved component.[3] For this purpose, we identify multiple clusters from the reference group by applying a Bayesian clustering model using finite mixture Gaussian distributions to the preference data of the reference group, Data Set I (Allenby, Arora, and Ginter 1998; Chung and Rao 2003). The averages of the residuals of each cluster for each product are regarded as the opinions of a virtual expert who represents the preferences of the customers in the corresponding group. Several dummy variables are defined to represent Y_{hj} data as reference consumer h's revealed or stated preference behavior (buying or rating) for product j, where h = 1, 2, ..., H and j = 1, 2, ..., J. The entry for Y_{hj} will range from 1 to R, and if there are available data on purchase behavior (buy or not buy), R = 2. The term PR_{hjr} is the probability of individual h's preference behavior (rating or buying) Y_{hj} for product j being r. We can model this as a finite mixture distribution that is a product of the probability of reference customer h's belonging to a certain class g, PG_{hg}, and the probability of h's preference behavior for product j by r, given h belongs to class g, denoted by $PR_{h|gjr}$:

$$(11) \quad PR_{hjr} = P(C_{h,r-1} < U_{hj} \le C_{hr}), \quad \text{where } U_{hj} = \beta_{0j} + \beta_{h0} + X^{'}_j\beta_{h1} + \varepsilon_{hj} = W_{hj} + \varepsilon_{hj},$$

[2]We thank the anonymous *JMR* reviewers for encouraging us to utilize the simultaneous approach for the estimation of models to improve model performance.

[3]It is also possible to use the residuals of the most similar consumer in the reference group as an expert without any model modification. However, we follow the cluster-based expert model because of the evidence in CF literature that a group of people in a cluster shows more reliable preference patterns than individual consumers in their corresponding groups (Breese, Heckerman, and Kadie 1998; Chien and George 1999). In addition, cluster-based expert models can respond to target customers' requests much more quickly than individual-based expert models as the firm obtains more target and reference customers.

$$= \sum_{g} P(h \in \text{cluster } g) P\Big(C_{h,r-1|h \in g} - W_{hj|h \in g} < \varepsilon_{hj|h \in g} \le C_{hr|h \in g} - W_{hj|h \in g}\Big)$$

$$= \sum_{g=1}^{G} PG_{hg} PR_{hjr|h \in g}.$$

The mixture distribution model includes the latent class and standard random coefficient choice models as special cases by allowing for consumer heterogeneity within and across clusters (Chung and Rao 2003). Web Appendix B (www.marketingpower.com/jmr_webappendix) describes the specification of prior distributions for model parameters in a Bayesian framework and the MCMC simulation for the identification of virtual experts (Model B). We estimate the posterior distributions of model parameters using reference consumers' rating data (Data Set 1) for posterior density estimation with a fixed number of classes.[4] The number of clusters is determined by choosing the model with the largest marginal likelihood, $m(y|M_G)$, for a mixture model M_G with G clusters. We calculate this marginal likelihood with a Bridge sampling estimator (Frühwirth-Schnatter 2004; Meng and Wong 1996), which provides robust estimates of marginal likelihoods for mixture models. Membership of each consumer in the reference group is determined by choosing the cluster with the largest posterior probability, $P(h \in g|\{y_{hj}\}_{j \in J(h)})$, where J(h) refers to a set of products rated/purchased by individual h:

$$(12)\quad P\Big(h \in g\Big|\{y_{hj}\}_{j \in J(h)}\Big) = \frac{P(h \in g) P\Big(\{y_{hj}\}_{j \in J(h)}\big|h \in g\Big)}{\sum_{g'}^{G}\Big[P(h \in g')\Big(P\{y_{hj}\}_{j \in J(h)}\big|h \in g'\Big)\Big]}.$$

For details on how to determine each consumer's membership, refer to Web Appendix B, Step 8 (www.marketingpower.com/jmr_webappendix).

We estimate the latent residuals of a standard attribute-based preference model for products with reference consumers' preference data in each cluster, using MCMC draws from the posterior distribution of latent residuals. We create a virtual expert to represent each cluster and assign group-specific statistics. These statistics include the mean and variance of residuals for each product as the opinion and precision level of the corresponding virtual experts for the same product.

In summary, our estimation and prediction procedure consists of four interrelated steps: (1) apply a standard preference model with a finite mixture of Gaussian distributions (Model B) to Data Set 1 to identify a number of clusters (referred to as virtual experts and indexed by g) in the reference group; (2) obtain the means and inversed variances of Bayesian latent residuals for reference consumers within each cluster, referred to as virtual experts' opinions and precision levels, respectively, for each product; (3) estimate Model A (the virtual expert model with virtual experts' opinions and precision levels as predictors in addition to product attributes) with the estimation data set (Data Set 2a); and (4) predict target consumers' preferences for products in Data Set 2b. In an actual situation, ratings in Data Sets 2a and 2b are the preference data the firm obtains from the target group and wants to predict, respectively.

EMPIRICAL APPLICATIONS

We now report two comprehensive applications of our approach and compare the results with other appropriate benchmark models. In the first application, we apply our model to recommendation data (stated preferences) obtained from the EachMovie database, which has become the best-known database for recommendation studies, and compare its performance with that of major CF models and attribute-based preference models. The second application involves viewing choices (revealed preferences) of movies.

Application 1: Online Preference Data for Movie Recommendations

To test the predictive performance of our model for recommendation services, we apply the model to the EachMovie database, the well-known benchmark data set for recommendation studies. The Compaq Systems Research Center provided this database by offering free Web-based recommendation services to people for 18 months up to September 1997. The database consists of 2,811,983 numeric ratings for 1628 movies (films and videos) entered by 72,916 users. This database contains (1) consumers' stated preference data (movie ratings with six-point ordered rating scales), (2) movie attributes (genres: action, animation, art and foreign, classic, comedy, drama, family, horror, romance, thriller), and (3) individual characteristics of consumers (age and gender). We first randomly sampled 300 movies for our empirical study. The distribution of movies by genre for this sample is similar to the genre distribution in the whole data set (see Web Appendix Table W1 at www.marketingpower.com/jmr_webappendix). We further randomly sampled 2000 people who rated at least three movies (at least two movies for model estimation and one movie for model validation) as target customers and 2335 people who rated more than five movies as reference customers. Including people who rated fewer than five movies did not improve the model performance in our pilot study. We used the movie ratings made by reference consumers (Data Set 1) in identifying virtual experts (for the calibration of Model B). We again randomly select two-thirds of the ratings made by each target customer for calibration of Model A (Data Set 2a) and the remaining one-third of the ratings for prediction of target customers' movie ratings (Data Set 2b). Web Appendix Table W1 provides some basic statistics for our sample on various descriptors in the data.

Model estimation. We applied the finite mixture model-based clustering to Data Set 1 consisting of the ratings of the reference group of 2335 people on 300 movies by varying the number of clusters up to 15, as described in Web Appendix B (www.marketingpower.com/jmr_webappen-

[4]Another optional method to identify clusters among the reference group is to use their latent residuals rather than their whole rating data. This method assumes that consumers' preference parameters for unobserved product attributes are homogeneous but those for observed attributes are not. We tested this optional approach by using a standard Gaussian mixture distribution. Its predictive performances were slightly worse in application 1 but nearly similar in application 2. Therefore, we believe this optional method can be used, depending on the nature of data. The empirical results of this optional method are available from the authors.

dix.[5] The log-marginal likelihood values for models with clusters varying in number from 2 to 15 varied from –44,012 (for the 2-cluster-based model) to –25,581 (for the 15-cluster-based model). The likelihood value monotonically increases up to –20,641 for 12 clusters and then decreases. Accordingly, we chose 12 clusters to identify virtual experts. Figure 2 provides the distribution of reference customers' individual memberships for the 12 clusters. Figure 3 shows the mean preference ratings by genre for virtual experts. It reveals that experts have divergent preferences in movies across genres.

We estimated the virtual expert models for the 2000 target consumers' preferences on 300 movies and experts' opinions and precision levels simultaneously using Data Sets 1 and 2a. As described in Web Appendix A (www.marketingpower.com/jmr_webappendix), we used MCMC simulation to estimate joint posterior distributions of model parameters for Models A and B simultaneously after incorporating the effects of nonignorable missing observations.[6] Tables 1 and 2 show the summary of model parameters for response choice behaviors. Table 3 shows the estimated model parameters for preference behaviors of the target group.

As Table 1 indicates, the reference group users are more likely to rate action, animation, and comedy movies than thriller, family, and drama movies. Table 3 shows that expert opinion makes a significant contribution in explain-

[5]We generated two Markov chains, in which we implemented 10,000 draws for the burn-in period and used 5000 draws from each of the two chains in addition for the estimation of posterior distributions.

[6]We generated two Markov chains, in which we implemented 30,000 draws of each chain for the burn-in period with convergence tests and used 10,000 draws (5000 draws from each of the two chains) for inference. In case experts have no opinions (no one in the corresponding cluster has not rated the corresponding movie), we set the corresponding mixing coefficients of the experts to 0, so the experts do not affect the predicted utility of the corresponding product for prediction.

Figure 2
THE MEMBERSHIP DISTRIBUTION OF THE REFERENCE GROUP FOR APPLICATION 1

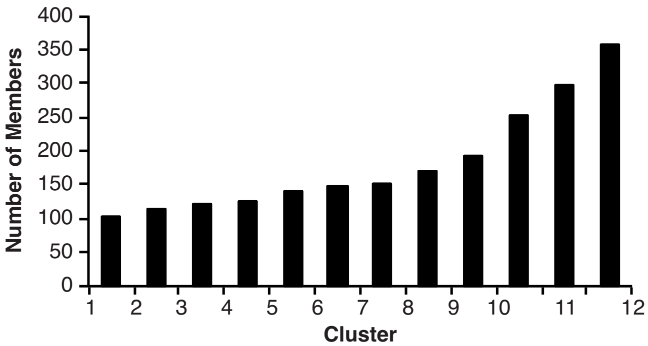
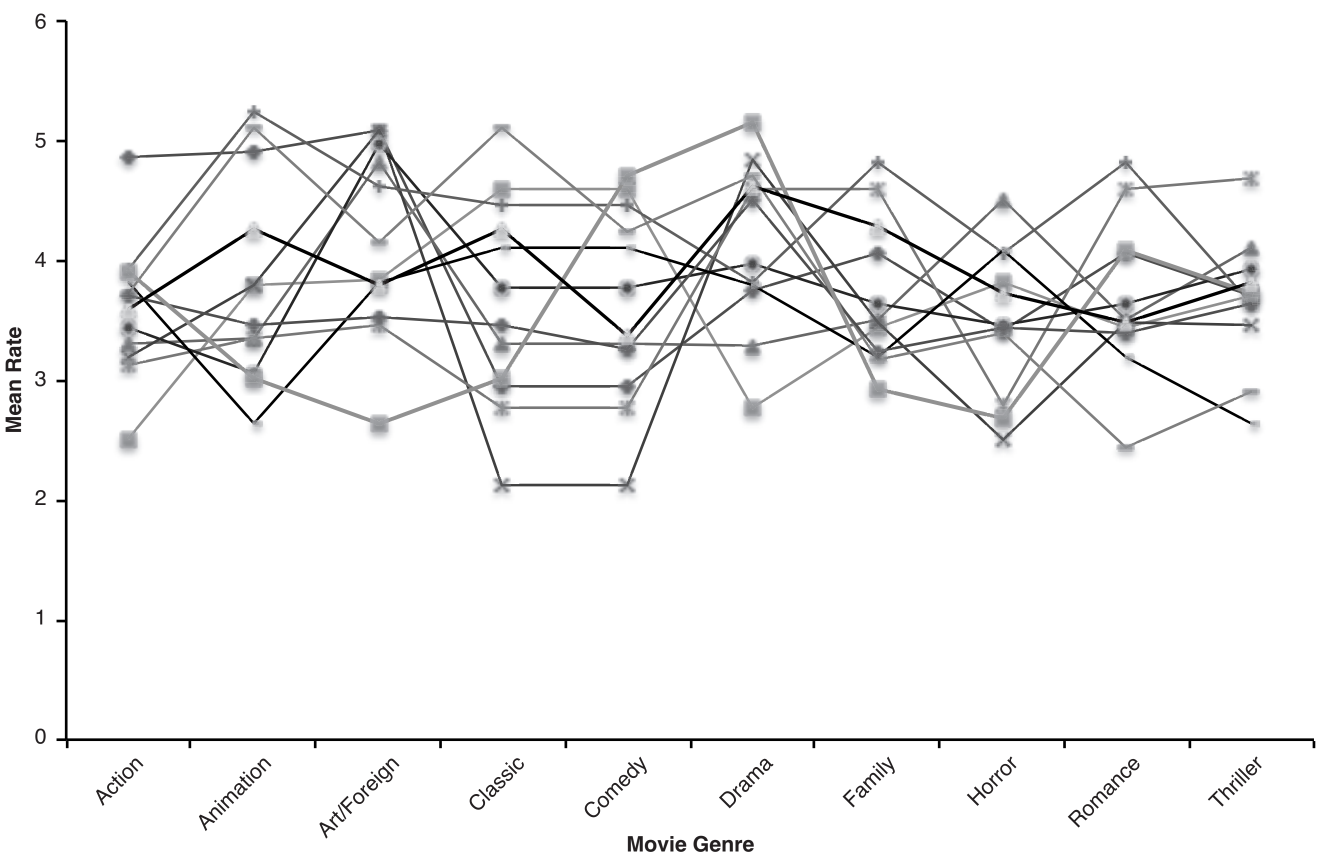


Figure 3
THE GENRE PREFERENCES OF VIRTUAL EXPERTS FOR APPLICATION 1



ing the missing parts of the utility of movies for target consumers, supporting the role of the virtual experts' opinions in our model. With regard to the grand mean values for the model parameters in the first-level equation, the coefficients of action, animation, comedy, horror, and romance genres are significant at the .01 level of Bayesian *p*-values. Not surprisingly, the estimated coefficients of experts are likely to be proportional to the size of the corresponding clusters. It implies that experts representing the large size of clusters are likely to make a relatively greater contribution to the improvement of the prediction accuracy of the model among experts.

Table 1
ESTIMATED MODEL PARAMETERS FOR RESPONSE CHOICE BEHAVIORS FOR APPLICATION 1

Variables	*Fixed Effect (SD[a])*	*Random Effect (SD Across Individuals[b])*	*Random Effect (SD Across Movies[c])*
Intercept	–2.892* (1.003[a])	.921	.502
Genre 1: Action	2.348* (.554)	.220	
Genre 2: Animation	2.015* (.629)	.510	
Genre 3: Art/Foreign	–.203 (.184)	.114	
Genre 4: Classic	.323 (.240)	.119	
Genre 5: Comedy	1.015* (.314)	.430	
Genre 6: Drama	–.348 (.352)	.211	
Genre 7: Family	–.381 (.144)	.121	
Genre 8: Horror	.529* (.200)	.204	
Genre 9: Romance	.944* (.481)	.360	
Genre 10: Thriller	–1.325 (.744)	.500	

[a]Average of the square roots of the diagonals of the corresponding covariance matrix draws obtained from MCMC.
[b]Average of standard deviations of MCMC draws for parameters across individual people.
[c]Average of standard deviations of MCMC draws for parameters across movies.
*Significant at the .01 level of Bayesian *p*-value.

Table 2
ESTIMATED CUTOFF POINTS FOR RESPONSE CHOICE BEHAVIORS FOR APPLICATION 1

Cutoff Points C	*Fixed Effect (SD[a])*	*Random Effect (SD Across Individuals[b])*
2	.227 (.164)	.133
3	.701* (.291)	.169
4	1.419* (.367)	.137
5	2.237* (.403)	.325

[a]Average of the square roots of the diagonals of the corresponding covariance matrix draws obtained from MCMC.
[b]Average of standard deviations of MCMC draws for parameters across individual people.
*Significant at the .01 level of Bayesian *p*-value.

Table 4 shows the relationship between the attribute (genre) coefficients and two demographic variables (age and gender). These coefficients seem to be face valid. For example, older consumers prefer classic, romance, and drama movies. Female consumers seem to prefer animation, art/foreign, family, and romance movies, while male consumers seem to prefer thriller and action movies.

Table 5 shows the relationship between the mixing coefficients and experts' three descriptor variables. We note that several of the estimated coefficients for the expert similarity variables are significant and positive, indicating that experts who are more similar to target customers in their opinions are more informative. A similar conclusion applies to the expert precision variable.

The cutoff points shown in Table 6 for the preference behavior reveal considerable variability in the thresholds for the higher preference ratings. Furthermore, the estimated correlation between response choices and preference behaviors is nonzero; this result underscores the importance of considering missing observations in the model.

Predictive validity test and model comparison. We test the predictive power of our model against that of three major CF models and four attribute-based preference models using a randomly selected one-third of each target customers' preference data for 300 movies (Data Set 2b).[7] In addition to details of the predictive power of each model, we provide a graphic diagnostic analysis to indicate the extent to which our model captures the unobserved components of product utilities compared with the other preference models.

[7]We estimate two versions of these three models with and without temporal effects by using additional data of timing of ratings.

Table 3
ESTIMATED MODEL PARAMETERS FOR PREFERENCE BEHAVIORS: ATTRIBUTE AND EXPERT OPINION COEFFICIENTS FOR LEVEL 1 FOR APPLICATION 1

Variables	*Fixed Effect (SD[a])*	*Random Effect (SD Across Individual People[b])*	*Variables*	*Fixed Effect (SD[a])*	*Random Effect (SD Across Movies[c])*
Intercept	–3.102* (1.336)	1.507	Intercept	—	.820
Genre: Action	.971* (.333)	.201	Expert opinion 1	.036 (.045)	.032
Genre: Animation	1.560* (.513)	.472	Expert opinion 2	.033 (.044)	.031
Genre: art/foreign	–1.537 (.801)	.694	Expert opinion 3	.039* (.026)	.021
Genre: Classic	.719 (.387)	.361	Expert opinion 4	.035* (.013)	.010
Genre: Comedy	.970* (.424)	.221	Expert opinion 5	.052 (.039)	.022
Genre: Drama	–.102 (.086)	.055	Expert opinion 6	.059* (.021)	.026
Genre: Family	–.271 (.238)	.148	Expert opinion 7	.071 (.052)	.044
Genre: Horror	–.462* (.212)	.244	Expert opinion 8	.083* (.032)	.024
Genre: Romance	.743* (.311)	.371	Expert opinion 9	.098 (.066)	.050
Genre: Thriller	–.735 (.445)	.413	Expert opinion 10	.103* (.047)	.056
			Expert opinion 11	.158* (.073)	.060
			Expert opinion 12	.232* (.103)	.071

[a]Average of the square roots of the diagonals of the corresponding covariance matrix draws obtained from MCMC.
[b]Average of standard deviations of MCMC draws for parameters across individual people.
[c]Average of standard deviations of MCMC draws for parameters across movies.
*Significant at the .01 level of Bayesian *p*-value.

Table 4
ESTIMATED RELATIONSHIPS BETWEEN GENRE COEFFICIENTS AND DEMOGRAPHICS FOR LEVEL 2 FOR APPLICATION 1

	Demographic Variables		
Genre Coefficient	*Age*[a]	*Gender*[b]	*Intercept*
Action	–.161 (.234)	–.211* (.101)	.311* (.172)
Animation	–.310* (.101)	.218* (.100)	.166 (.542)
Art/foreign	.224 (.143)	.122* (.044)	–.423* (.132)
Classic	.476* (.201)	.398 (.223)	–.717* (.330)
Comedy	–.214* (.100)	.555 (.391)	.101 (.164)
Drama	.255* (.101)	–.106 (.099)	–.162 (.100)
Family	–.197 (.122)	.402 (.300)	–.330 (.215)
Horror	.146 (.102)	–.211 (.161)	–.284* (.118)
Romance	.411* (.192)	.276* (.099)	–.198 (.150)
Thriller	.103 (.078)	–.623* (.228)	.401 (.306)

[a]Age variable is standardized.
[b]Gender variable is coded as 1 = female and 0 = male.
*Significant at the .01 level of Bayesian *p*-value.
Notes: The value in each parenthesis is the corresponding standard deviation.

Table 5
ESTIMATED RELATIONSHIP BETWEEN MIXING COEFFICIENTS AND EXPERT INFORMATIVENESS VARIABLES FOR APPLICATION 1

Expert Opinion	*Expert Similarity*	*Expert Precision*	*Intercept*
Expert 1	.367 (.255)	.283* (.103)	—**
Expert 2	.397* (.093)	.333 (.204)	–.532* (.153)
Expert 3	–.098 (.101)	.344* (.121)	–.777* (.183)
Expert 4	–.116* (.037)	.402* (.199)	–.805 (.645)
Expert 5	.313* (.128)	.331 (.220)	–.527 (.277)
Expert 6	–.083 (.090)	.235 (.166)	–.331 (.205)
Expert 7	.412* (.163)	.332 (.224)	–.523* (.243)
Expert 8	.367 (.197)	.340* (.140)	–.790 (.554)
Expert 9	.511 (.377)	.226* (.100)	–.495 (.451)
Expert 10	.469* (.137)	–.233* (.105)	–.279 (.261)
Expert 11	.401* (.167)	.242 (.137)	–.50* (.148)
Expert 12	.477* (.154)	–.142 (.110)	.131 (.099)

*Significant at the .01 level of Bayesian *p*-value
**This intercept parameter is set to zero for model identification.

First, we discuss the reasons for choosing the specific models for comparison and present a summary table showing the aspects they do and do not consider. We include the baseline predictor model that considers the independent effects of consumers and products only (Model 1) (Bell, Koren, and Volinsky 2008; Takács, Pilászy, and Németh 2008). We select two major CF models that are among the best candidates of CF models used by Bellkor, the Netflix competition winning team (Bell, Koren, and Volinsky 2008; Koren 2009)[8]: the neighborhood model (Model 2) and the SVD++ (Model 3), which is an MF model. The SVD++ is a variant of SVD (singular value decomposition) models, which outperformed other variants of MF models (Bell, Koren, and Volinsky 2008; Koren 2009). Matrix factorization models, including the SVD++, decompose the rating matrix as a product of two lower-dimensional matrixes P (product-specific) and Q (consumer-specific) using singular value decomposition to obtain a rank-k approximation of the original matrix. From our modeling perspective, we can interpret this approach as estimating a low-dimensional characteristic vector for each user and a low-dimensional product attribute vector for each movie.

We consider four attribute-based preference models (Models 4, 5, 6, and 7). Model 4 is a standard consumer preference model with a unimodal consumer heterogeneity distribution estimated using an ordinal probit framework. Model 5 is the same as Model 4 but with multimodal consumer heterogeneity specified through a finite mixture of Gaussian distributions. Indeed, this model can be used as a clustering-based recommendation model. Model 6 is a hierarchical linear regression model that considers the effects of product attributes and also product heterogeneity for the unobserved component (Ansari, Essegaier, and Kohli 2000). Model 7 is an extension of Model 6 that also considers the effects of response choice behaviors in an ordinal probit framework (Ying, Feinberg, and Wedel 2006). Moreover, we test two versions of our virtual expert model: a model with only the most similar expert (Model 8) and a model with all 12 experts (Model 9). We summarize the mathematical descriptions of product utilities for these models in Table 7. Web Appendix C (www.marketingpower.com/jmr_webappendix) provides more details on what information is used and which effects are considered in what way by each model in a comparable and consistent manner and, in particular, shows how the unobserved component is handled either implicitly or explicitly in these models compared with our model.

We used Data Sets 1 and 2a for the estimation of all models (Models 1–9), and we used Data Set 2b for their prediction tests. Table 8 shows the predictive performances of models using the criteria of hit rates and root mean square error (RMSE) values. Not surprisingly, Model 1 (the baseline predictor model) is the worst performer among all models. Model 3 (SVD++) with temporal effect shows the best performance among CF models and the attribute-based

[8]The Netflix competition winning team's full model is an ensemble of prediction values from hundreds of independent models (Koren 2009). However, most of their models are variants of underlying major methods and the winning team acknowledged that most performance of their model is achievable by neighborhood and MF methods (Bell, Koren, and Volinsky 2008; Takács, Pilászy, and Németh 2008).

Table 6
ESTIMATED CUTOFF POINTS FOR APPLICATION 1

Cutoff Points C	*Fixed Effect (SD*[a]*)*	*Random Effect (SD Across Individuals*[b]*)*
2	.301* (.131)	.056
3	.788 (.316)	.122
4	1.487* (.617)	.406
5	2.499* (1.137)	.669
Correlation[c]	.092* (.044)	.041

[a]Average of the square roots of the diagonals of the corresponding covariance matrix draws obtained from MCMC.
[b]Average of standard deviations of MCMC draws for parameters across consumers.
[c]This is the correlation between preference and response choice.
*Significant at the .01 level of Bayesian *p*-value.

Table 7

MODEL STRUCTURE COMPARISON*

Model	*Utility Structure* $Y_{ij} = r \text{ if } C_{i,r-1} < U_{ij} \leq C_{ir} \quad r = 1, 2, \ldots, R$ $U_{ij} = \beta_{i0} + \beta_{0j} + X'_j\beta_i + X^*_j\beta^*_i$
Model 1: Baseline model Linear regression (Bell, Koren, and Volinsky 2008)	$E(y_{ij}) = b_0 + \beta_{i0}(t_{ij}) + \beta_{0j}(t_{ij})$
Model 2: Neighborhood model Linear regression (Koren 2009)	$E(y_{ij}) = b0 + \beta_{i0}(t_{ij}) + \beta_{0j}(t_{ij})\kappa(t_{ij}) + \lvert R(i)\rvert^{-1/2} \sum_{j' \in R(i)} \left[y_{ij} - \left(b_0 + b_{i0} + b_{j0}\right)\right]\beta^1_{jj'} + \lvert N(i)\rvert^{-1/2} \sum_{j' \in N(i)} \beta^2_{jj'}$
Model 3: Matrix factorization –SVD++ Linear regression (Bell, Koren, and Volinsky 2008; Koren 2009)	$E(y_{ij}) = b_0 + \beta_{i0}(t_{ij}) + \beta_{0j}(t_{ij})\kappa_i(t_{ij}) + q'_j\left(p_i(t_{ij}) + \lvert N(i)\rvert^{-1/2} \sum_{j' \in N(i)} \beta^3_{j'}\right)$
Model 4: Preference model with unimodal consumer heterogeneity	Ordinal probit $E(U_{ij}) = \beta_{i0} + X_j\beta_{i1}$ $\beta_i \sim N(\mu, \Sigma)$
Model 5: Preference model with multimodal consumer heterogeneity	Ordinal probit $E(U_{ij}) = \beta_{i0} + X'_j\beta_{i1}$ $\beta_i \sim N(\mu_g, \Sigma_g), \text{ if } i \in \text{cluster } g$
Model 6: Linear regression (Ansari, Essegaier, and Kohli 2000)	$E(Y_{ij}) = \beta_{i0} + \beta_{0j} + X'_j\beta^x_i + z'_i\beta^z_j$
Model 7: Ordinal probit regression (Ying, Feinberg, and Wedel 2006)	$E(U_{ij}) = \beta_{i0} + \beta_{0j} + X'_j\beta^x_{i1} + z'_i\beta^z_{1j} + \rho_i \frac{\varphi(\beta^s_i X_j)}{\Phi(\beta^s_i X_j)}$
Models 8 and 9: Virtual expert model	Ordinal probit $E(U_{ij}) = \beta_{0j} + \beta_{i0} + X'_j\beta_{i1} + \eta'_j\alpha_{.ij} + \rho_i \frac{\varphi(\beta^s_i X_j)}{\Phi(\beta^s_i X_j)}$

*Refer to Web Appendix C (www.marketingpower.com/jmr_webappendix) for more details on model specifications and notation.

models (4–7). Model 2 (neighborhood model) is less accurate than Model 3; this is consistent with the previous studies (Bell, Koren, and Slovensky 2008). Given that available information on product attributes is limited only to movie genre, it is not surprising that the CF models that do not rely on the attribute information showed better predictions than the four attribute-based models.

The virtual expert model with multiple experts (Model 9) outperforms all other models. Even the virtual expert model with a single expert (Model 8) significantly outperforms Models 1–7, except SVD++ (Model 3), in both in- and out-of-sample fits. Model 3 (with additional information on date of rating) performs as well as Model 8. The virtual expert model outperforms the CF models (Models 1–3), mainly because it incorporates product attributes and consumer characteristics and residuals from preferences of others with a more appropriate modeling framework. The virtual expert model outperforms choice models (Models 4 and 5), which implies that it improves these attribute-based models by capturing the unobserved component of a product utility as a function of other consumers' latent residuals. The comparison of the virtual model with Model 5 (the standard preference model with multimodal consumer heterogeneity) shows that experts' opinions contribute much to the model prediction. Finally, the results show that the virtual expert model (Model 8 or 9) has some advantages over Models 6 and 7 that capture the nonquantifiable attributes with the interaction term between consumer demographics and product-specific parameters. In our view, the interaction effect with age and gender in Models 6 and 7 can only partially capture the unobserved component, while the virtual expert model almost fully captures the unobserved component.

To provide some insights into the conditions in which the virtual expert model will perform well against other mod-

Table 8
MODEL COMPARISON FOR APPLICATION 1

Model	*Description of Information Used for Prediction*[a] *I*	*II*	*III*	*IV*	*Hit Rate (RMSE)* *In-Sample*	*Holdout Sample*
CF Models						
1. Baseline model					.30/.37[a] (1.82/1.62[a])	.27/.33[a] (2.02/1.80[a])
2. Neighborhood model	✓		✓	✓	.43/.46[a] (1.39/1.30[a])	.38/.40[a] (1.59/1.50[a])
3. MF model (SVD++)	✓		✓	✓	.46/.49[a] (1.27/1.19[a])	.43/.45[a] (1.40/1.30[a])
Attribute-Based Preference Models						
4. (Preference model with unimodal consumer heterogeneity)		✓			.39 (1.62)	.33 (1.75)
5. (Preference model with multimodal consumer heterogeneity-clustering)		✓			.42 (1.43)	.37 (1.72)
6. (Linear regression with consumer and product heterogeneity)	✓	✓	✓		.45 (1.30)	.40 (1.51)
7. (Preference model with missing data and product and consumer heterogeneities)	✓	✓	✓		.47 (1.23)	.43 (1.43)
Hybrid Modeling Approach						
8. (Virtual expert model with the most similar expert)	✓	✓	✓	✓	.49 (1.19)	.45 (1.30)
9. (Virtual expert model with multiple experts)	✓	✓	✓	✓	.53 (1.01)	.49 (1.18)

[a]The two entries represent the performance measures without and with the effect of timing of ratings.
I: Nonignorable missing data (response choice behavior): ✓ = considered; blank = not considered.
II: Product attributes: ✓ = present; blank = absent.
III: Scale heterogeneity: ✓ = present; blank = absent.
IV: Use of others' ratings (preference similarity): ✓ = present; blank = absent.

els, we conduct three additional sensitivity analyses for studying three factors: (1) the effect of consumer heterogeneity using a graphic analysis of latent residuals, (2) the effect of the number of ratings by each target customer on the predictive power of major models, and (3) the effect of the number of ratings by each reference customer on the predictive power of the virtual expert model.

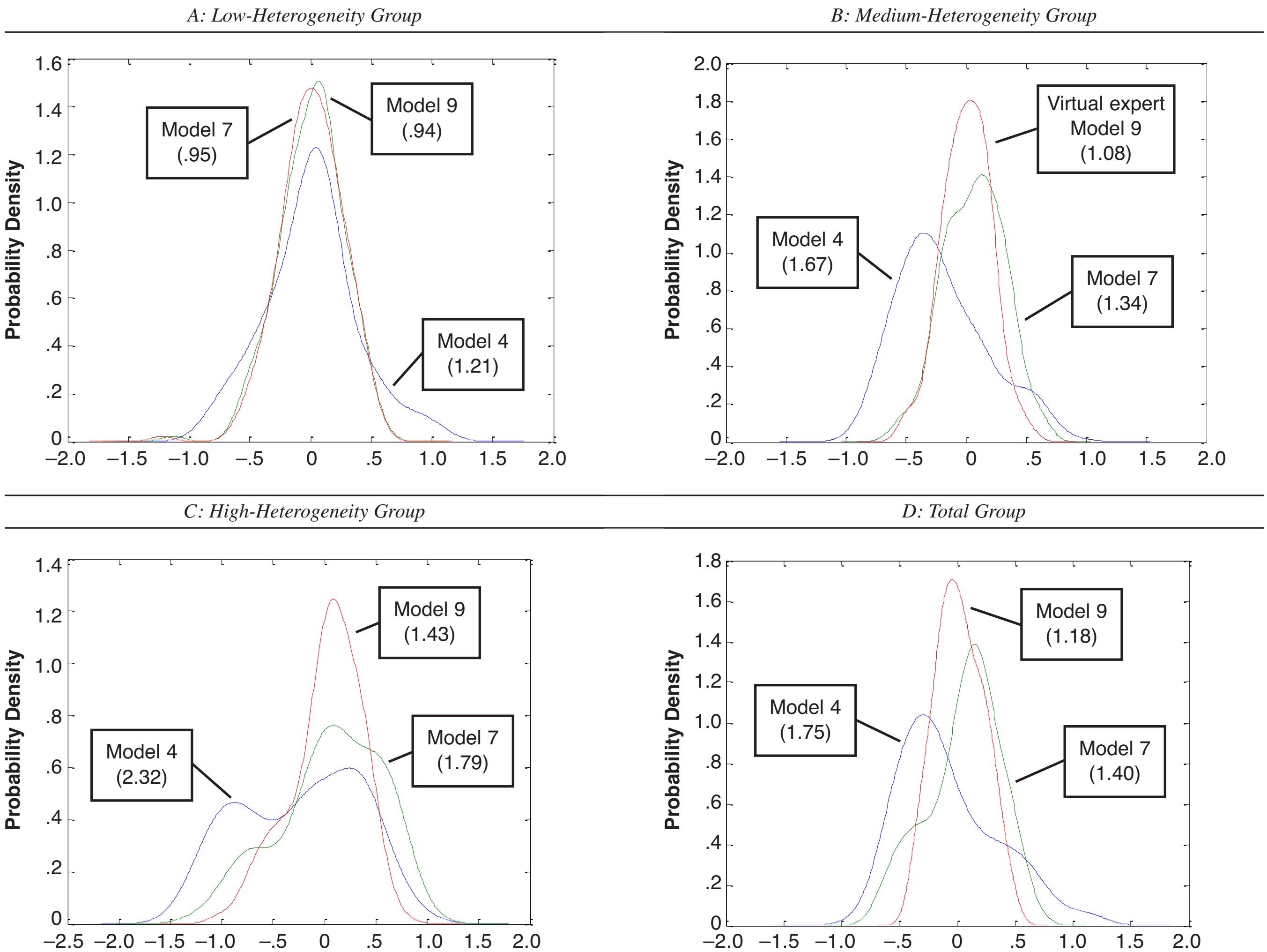
The effect of consumer heterogeneity. We visually compare the distributions of latent residuals obtained from three consumer preference models (Models 4, 7, and 9) to provide an understanding of how well preference models capture the unobserved component for three groups of target consumers: (1) low-heterogeneity group, (2) medium-heterogeneity group, and (3) high-heterogeneity group. For this purpose, we first compute the mean absolute deviance (MAD) between the target consumer's ratings. Then, we sample three groups: high-homogeneity group with the lowest MADs (20% of target consumers), medium-homogeneity group in the middle MAD values (approximately 40%–60% of target consumers), and low-homogeneity group with high values of MAD (20% of target consumers). Figure 4 shows the residual distributions of three preference models. Panels A, B, and C show that the more heterogeneous the target consumer is, the more likely that the distribution of unobserved components of standard preference models (e.g., Model 4) is multimodal.

In the case of the low-heterogeneity group, the residual distributions of three models in Figure 4, Panel A, are similar, with low standard deviations (less than .4). This implies that the three preference models performed well for the predictions of homogeneous consumers' preferences. However, the gaps among model performances in terms of RMSEs become larger in the case of relatively heterogeneous consumers, as shown in Figure 4, Panels B and C. This implies that Model 4 (standard preference model) becomes less accurate for the predictions of heterogeneous consumers and that Models 7 and 9, which capture the unobserved components, become more accurate than a standard preference model. In addition, Model 7 does not sufficiently capture the unobserved component compared with the virtual expert model, especially when consumers are heterogeneous in their preferences. The latent residual distributions of our model are tighter and more unimodal for the virtual expert model (Model 9) than those of the other two preference models. This implies that the virtual expert models capture the unobserved components well and also handle multimodality of latent residuals. We observe similar results when comparing latent residuals of three models obtained from all consumers, as shown in Figure 4, Panel D. This additional diagnostic analysis shows another way of utilizing one of our by-products, the latent residuals, and enables researchers to visually diagnose the drawbacks of their models.

The effect of the amount of target customers' preference data. We grouped target consumers into five subgroups according to their number of ratings, sometimes called users' support, in Data Set 2a and tested the effect of the number of ratings by each target customer on the predictive powers of major models in each area: the neighborhood model (Model 2), the MF model (Model 3), the attribute-based preference models (Models 4 and 7), and the virtual

Figure 4

GRAPHICAL DISTRIBUTIONS OF THE LATENT RESIDUALS FOR APPLICATION 1



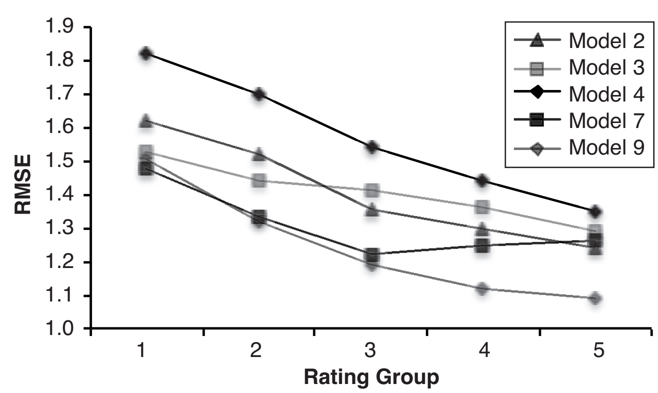
Notes: The value in each parenthesis is RMSE.

expert model with multiple experts (Model 9); Figure 5 presents the results.

In the first group (the number of ratings per person < 6), the MF model (Model 3) was the best and the neighborhood model was the worst in terms of RMSE. Because the preference data are limited for this group, the performance of our model became worse but was not as bad as the neighborhood model because the model included product attributes and multiple experts. The performance of all models improved as the number of consumer preference ratings increased. Notably, MF models performed well when user support is between 11 and 20, but their performance did not improve significantly as consumers provide more ratings; in contrast, Models 7 and 9 performed better with a larger number of preference ratings. This comparison shows that the MF model is still useful for early users with a limited number of preference ratings. Our model can make a greater contribution as the recommendation system accumulates user preference ratings. This result implies that if a firm employs the ensemble strategy of combining multiple mod-

Figure 5

THE EFFECT OF THE NUMBER OF RATINGS OF TARGET CUSTOMERS ON MODEL PERFORMANCE FOR APPLICATION



Notes: Group 1 = Less than 6 ratings; Group 2 = 6–10 ratings; Group 3 = 11–20 ratings; Group 4 = 21–40 ratings; and Group 5 = 41 ratings or more.

els, including both the MF model and our model can be beneficial.

The effect of the amount of reference customers' preference data. It is well known that most recommendation databases contain many raters who rated only a few movies. Collaborative filtering studies using clustering (Bell, Koren, and Volinsky 2008) report that using people who rated two or three times for clustering is likely to generate unreliable clusters. In general, the more products the reference customers have experience with (i.e., the more products they have rated), the better will be the contribution of experts' opinions to the performance of the model. To shed some light on this, we conducted a sensitivity analysis. We randomly sampled 3000 customers who rated at least one product and developed five smaller data sets (A, B, C, D, and E) of people who rated more than one, three, five, seven, and nine products, respectively. Table 9 shows the performances of the model for these data sets.

The empirical analysis reported in Table 8 uses Data Set C, consisting of customers who rated more than five products. It outperformed the other data sets in terms of in-sample (hit rates and log-likelihoods) and out-of-sample fits. Thus, we find that including less experienced customers for the reference group lowers the accuracy of the model while generating more clusters. Including those people is likely to overevaluate the informativeness of experts (less accurate estimates of membership probabilities and overevaluated precision level). The result implies that using as large a sample as possible (subject to computational feasibility) is preferable in a real situation; however, researchers should exclude less experienced people (e.g., those who rated fewer than five products).

Application 2: Offline Purchase Data for Movie Choice Behaviors

Although we have shown the superiority of our approach against other major models for recommendation algorithms, it is important to demonstrate that our modeling approach is a general method to improve any preference model for experience products or products with too many attributes relative to the available amount of preference data. To accomplish this objective, we now apply our model to revealed preference data collected through personal interviews on viewing behavior (or revealed preferences) for 30 movies (15 foreign and 15 Korean) from a convenience sample of 312 consumers in the Seoul metropolitan area. The movie list appears in Web Appendix Table W2 (www.marketingpower.com/jmr_webappendix). In addition to viewing behavior (whether the respondent saw the movie), we collected data on movie attributes (genre and national origin) and individual characteristics (age and gender). We classified movie genre into seven categories (drama, romance, action, animation, comedy, horror, and fantasy), following those of the Korean Film Council; and national origin into three categories (Western, Asian except Korean, and Korean). Because this data set had no missing data, we do not consider response choice behavior. We used a nested model with R = 2 (probit model).

Model estimation and prediction. We obtained three data sets (Data Sets 1, 2a, and 2b) by randomly dividing respondents into two groups of 200 and 112 people and assigning them to the reference consumer group (Data Set 1) and the target consumer group (Data Set 2), respectively. In addition, we randomly allocated two-thirds of the ratings of each target customer in Data Set 2 for model estimation (Data Set 2a) and one-third of the ratings of each target customer for prediction (Data Set 2b). We identified three virtual experts (clusters) from Data Set 1 by clustering reference consumers for two to five clusters and using the highest marginal likelihood value as compared with the other solutions. The values of the marginal likelihood, calculated with a Bridge sampling estimator (Frühwirth-Schnatter 2004; Meng and Wong 1996), are –365.1, –333.9, –353.4, and –359.0, respectively, for two, three, four, and five clusters. The membership probabilities were 26%, 34%, and 40%, respectively, for the three clusters in this solution; these are numbered 1–3. We estimated virtual experts' opinions (residuals) and the virtual expert model with Data Sets 1 and 2a. We also predicted 112 target consumers' choice behaviors regarding 30 movies with Data Set 2b, using information on product attributes, individual characteristics, and virtual experts' opinions and confidences.[9] The parameter estimates, such as grand means and standard deviations of the estimated parameters averaged over consumers, are summarized in Web Appendix Tables W4, W5, and W6 (www.marketingpower.com/jmr_webappendix).

For the predictive validity test, we compared the performances of our two models (Models 4 and 5) with Models 1, 2, and 3 applied to Data Set 2b (see Table 10). Model 1 is a standard choice model with product attributes only, and Model 2 is another standard choice model with product attributes, consumer characteristics, and unimodal consumer heterogeneity. Model 3 is equivalent to Model 2 but is based on multimodal consumer heterogeneity. In addition to product attributes and consumer characteristics, Model 4

[9]We used MCMC simulation for the estimation of joint posterior distributions of model parameters with two Markov chains. We implemented 5000 draws of each chain for the burn-in period with convergence tests and used 3000 draws from each chain for inference.

Table 9

SENSITIVITY OF MODEL PERFORMANCE FOR VARYING REFERENCE GROUP SIZES FOR APPLICATION 1

	Products Rated More Than	*Number of Reference Customers*	*Number of Clusters Identified*	*In-Sample Fit (Hit Rate/ Log-Marginal Likelihood)*	*Out-of Sample Fit Data Set (Hit Rate)*
A	1	3000	14	.483 / –27276.1	.478
B	3	2702	13	.501 / –2679.3	.486
C	5	2335	12	.524 / –26122.4	.494
D	7	2101	12	.511 / –26576.6	.484
E	9	1721	11	.485 / –27299.6	.476

Table 10

MODEL PERFORMANCE FOR APPLICATION 2

	In-Sample Fits		*Out-of-Sample Fit*
Model	*Log-Marginal Likelihood*	*Hit Rate (In-Sample)*	*Hit Rate (Out-of-Sample)*
Standard Choice Model			
1. Choice model with product attributes only and unimodal consumer heterogeneity	–955.25	76%	70%
2. Choice model with product attributes, consumer characteristics, and unimodal consumer heterogeneity	–934.77	78%	73%
3. Choice model with product attributes, consumer characteristics, and multimodal consumer heterogeneity	–903.48	80%	74%
Virtual Expert Model			
4. Choice model with the most similar expert	–801.42	86%	83%
5. Choice model with multiple experts	–758.91	89%	88%

is a virtual expert model that uses the most similar experts' opinions only, whereas Model 5 is another virtual expert model that uses all three experts.

Model 2 outperformed Model 1 in both in- and out-of-sample fits, implying that respondents' characteristics are important variables for explaining consumer heterogeneity and for predicting their choice behaviors. Model 3 outperformed the other standard preference models (Models 1 and 2) in both in- and out-of-sample fits, implying that multimodal consumer heterogeneity is more helpful for predicting their choice behaviors than unimodal consumer heterogeneity. Virtual expert models (Models 4 and 5) outperformed standard choice models with no exception in terms of in- and out-of-sample fits. This means that virtual expert models successfully used additional information extracted from the Bayesian residuals. The best model is Model 5, which uses multiple experts' opinions with mixing coefficients adjusted for each expert with the help of two informative variables. Compared with Model 3, the out-of-sample fit of that model was improved by 18.9% by including virtual experts' opinions.

DISCUSSION AND FURTHER RESEARCH

We develop a general methodology for preference models for experience products or products with a large number of product attributes that takes into account the effects of unobservable product attributes. We apply our virtual expert model to two situations of stated preference (ratings) and revealed preference (choice) data. In both applications, we show considerable improvements in predictive power with the virtual expert model. This study makes four methodological contributions to the literature on choice modeling.

First, we develop a general methodology that can be used for the improvement of the predictive power of any preference model. This is true regardless of whether they are stated preference data obtained from Internet recommendation systems or revealed preference data from actual purchase data, especially for experience products with non-quantifiable (or unobservable) product attributes. Second, our model links two methodological streams—CF and consumer choice models—both conceptually and methodologically. If this model is used for recommendations in an actual situation, it can generate a new type of word-of-mouth effect without direct communication among consumers. Our model can be interpreted as a model that embeds automated word-of-mouth information (which is the primary fuel of the predictive power of CF) into the random utility model in a complementary manner.

Third, our model enables firms to fully use all different types of information (e.g., product attributes, consumer characteristics, consumer preference similarity) available in e-retailers' consumer databases. In addition, it also allows firms to use both the revealed and stated preference data simultaneously. Note that most online shopping malls, such as Amazon.com and CDnow.com, now routinely collect data on consumers' stated preference data (ratings) on their first visits and augment them by revealed preference data (actual purchases) over time. Fourth, the most important value of our model is that our method can be embedded into any preference model or any model that generates residuals from methodological perspective. For example, including the virtual expert model into the MF model can improve the performance of recommendation systems.

The simultaneous estimation approach is the better estimation strategy for general choice studies mainly focusing on the improvement of prediction accuracy. Nevertheless, we should note one important limitation of our model. The simultaneous estimation approach adopted in this study for model estimation may significantly restrict the applicability of the model for online recommendation providers, because it requires the firm to estimate the parameters for Model B with the same data (Data Set 1) repetitively whenever the firm obtains a new group of target customers or whenever its goal is to provide better recommendations to the same group of target consumers at a later time. Given that Data Set 1 for the estimation of Model B is likely to consist of a relatively large amount of preference data in the case of online shopping malls, its repetitive computational burden may undermine the applicability of our model. Therefore, we suggest the two-stage estimation approach as a practical option a firm can choose. We believe the firms accessing a huge amount of data from many customers, such as those using online shopping malls, are likely to prefer the two-stage estimation approach, because the estimates of the opinions and precision levels of virtual experts obtained from the application of Model B to reference customers' data only are likely to be less efficient but not that much worse, because the data set usually contains a great deal of preference data from many reference customers. Firms interested in more efficient parameter estimates for prediction accuracy or choice modelers can use the simultaneous estimation approach, while firms interested in reducing time to respond to target customers can use the two-stage estimation.

In addition, because our study requires reserving people for reference groups, this could be a drawback for its application to studies with a small number of observations and accompanying predictive testing. However, use of virtual experts' opinions (estimated from the reference group data) improves the prediction power of our model. This is a matter of which model uses information more efficiently as long as target customers are known a priori, as in the case of online recommendations. The empirical results show that our approach utilizes such information more efficiently than the other preference models compared.

Regarding the size of the target group, this is essentially predecided by the flow of the inbound customers who visit the company's website. We suggest that using as large a sample for the reference group as possible according to computational feasibility will be preferred if a firm needs to use this model repeatedly with a different number of new target customers over time; however, researchers should exclude less experienced people (e.g., those who rated less than five products).

In conclusion, we identify some directions for further research. First, it is possible to improve the predictive accuracy of the virtual expert model, especially for new users, by using both the typical preference data and other types of behavioral data obtained from product-specific consumption or web searches for shopping. Second, to enhance the scalability of the model for recommendation services, it is desirable to simplify the model estimation procedures or to extend the model by incorporating adaptive learning (Chung, Rust, and Wedel 2009). Third, a good future research topic would be to improve the virtual expert model with the ideas suggested by other studies on consumers' preference similarities. For example, preference interdependence models such as Yang and Allenby's (2003) basically augment utilities with a correlated error structure across consumers for the unobserved component, similar to our model. Instead of preference similarity, they use geographic similarity to explain how a consumer's utility is influenced by others. Although their model addresses a different problem and uses a different type of data, if additional information on geographic similarities among customers is available, it would be worthwhile to compare those models with ours and to find a way of combining those ideas to enhance our model.[10]

APPENDIX: SUMMARY OF THE VIRTUAL EXPERT MODEL

Observed Preference Data: Purchase or Rating

$$Y_{ij} = r \quad \text{if } C_{i,r-1} < U_{ij}^{T} \le C_{ir} \quad r = 1, 2, \ldots, R, \quad \forall \text{ i and j.}$$

Latent Utility for Preference Behaviors

$$\text{(Form A) } U_{ij} = W_{ij} + V_{ij} + \varepsilon_{ij}, \text{ where } \varepsilon_{ij} \sim N(0, 1) \; \forall \text{ i and j.}$$

[10]We thank one of the anonymous *JMR* reviewers for clarifying the similarities and dissimilarities of another model relative to ours.

Observed Component: W_{ij}

$$W_{ij} = \beta_{0j} + \beta_{i0} + X_j'\beta_{i1} \text{ for all i and j,}$$
$$\text{where } X_j' = (X_{j,\,\text{genre1}}, X_{j,\,\text{genre2}}, \ldots, X_{j,\,\text{genre P}}).$$

$$\beta_{0j} \sim N(0, \sigma_\beta) \text{ for all j and } (\beta_{i0}, \beta_{i1}) \sim MVN(\mu_{\beta_i}, \Sigma_\beta) \text{ for all i.}$$

$$\mu_{\beta_i} = \Psi' Z_i, \text{ where } Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iD})',$$
$$Z_{iD} = 1 \text{ (consumer characteristics)}, \Psi_i = (\psi_1, \psi_2, \ldots, \psi_P),$$
$$\text{and } \psi_p = (\psi_{p1}, \psi_{p2}, \ldots, \psi_{pD})'.$$

Here, P is the number of product attributes, and D is the number of individual characteristics with an intercept.

Unobserved Component: V_{ij}

With $\alpha_{ij} = (\alpha_{1ij}, \ldots, \alpha_{gij}, \ldots, \alpha_{Gij})$ denoting mixing coefficients and $\eta_j' = (\eta_{1j}, \ldots, \eta_{gj}, \ldots, \eta_{Gj})$ denoting the estimates of latent residuals of product j for virtual expert g obtained from the standard attribute-based model,

$$V_{ij} = \eta_j'\alpha_{ij}, \text{ s.t. } \sum_g \alpha_{gij} = 1 \text{ and } 1 \ge \alpha_{gij} \ge 0.$$

Furthermore, mixing coefficients are determined as follows:

$$\alpha_{gij} = \frac{\exp\left(\gamma_g'\varpi_{gij}\right)}{\sum_{g'} \exp\left(\gamma_{g'}'\varpi_{g'ij}\right)} \text{ for all i, g, and j,}$$

where $\varpi_{gij} = (1, \varpi_{1gi}, \varpi_{2gi})$ denotes an intercept, preference similarity between target customer i and expert g (membership probability), and expert g′ precision (inversed variance of residuals) levels. $\gamma_g' = (\gamma_{g0}, \gamma_{g1}, \gamma_{g2})$ denotes the corresponding parameters for ϖ_{gij}.

The Effects of Response Selection Behaviors

If it is necessary to include the effects of nonignorable missing observations due to consumers' response selection behaviors, the utility model should be reformulated as follows, and the response probabilities should be conditioned on response selection behaviors as follows.

$$\text{(Form B)} \quad E(U_{ij}) = W_{ij} + V_{ij} + \rho_i \frac{\varphi\left(W_{ij}^s\right)}{\Phi\left(W_{ij}^s\right)},$$
$$\text{where } \tanh^{-1}(\rho_i) \sim N\left(\mu_\rho, \sigma_\rho\right).$$

$$W_{ij}^s = \beta_{0j}^s + \beta_{i0}^s + X_j'\beta_i^s, \text{ (the utility of target i's responding to product j for preference evalution)}$$

$$\beta_i^s = \Psi^{s'} Z_i + \xi_i^s, \quad \xi_i^s \sim N(0, \Sigma_\beta^s) \quad \text{for all i and j,}$$

$$\text{where } \Psi^s = (\psi_1^{s'}, \psi_2^{s'}, \ldots, \psi_P^{s'}), \psi_p^s = (\psi_{p1}^s, \psi_{p2}^s, \ldots, \psi_{pD}^s)',$$
$$\text{and } \xi_i^s = (\xi_{i1}^s, \xi_{i2}^s, \ldots, \xi_{iP}^s).$$

The corresponding priors for the response choice model are the same as those for the preference choice models. If the model considers the effects of missing observations, the response choice and the preference choice models are simultaneously estimated using the following bivariate normal distribution straightforwardly: $(\varepsilon_{ij}, \varepsilon^{s}_{ij}) \sim MVN(1, 1, 0, 0, \rho_i)$.

REFERENCES

Albert, James and Siddhartha Chib (1995), "Bayesian Residual Analysis for Binary Response Regression Models," *Biometrika*, 82 (4), 747–59.

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (August), 384–89.

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommender Systems," *Journal of Marketing Research*, 37 (August), 363–75.

Bell, Robert and Yehuda Koren (2007), "Improved Neighborhood-Based Collaborative Filtering," in *Proceedings of KDD Cup Workshop at SIGKDD'07, 13th ACM International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 7–14.

——, ——, and C. Volinsky (2008), "The BellKor 2008 Solution to the Netflix Prize," technical report, AT&T Labs Research, (accessed January 30, 2012), [available at http://www2.research.att.com/~volinsky/netflix/Bellkor2008.pdf].

Bodapati, Anand V. (2008), "Recommendation Systems with Purchase Data," *Journal of Marketing Research*, 45 (August), 77–93.

Breese, John S., David Heckerman, and Carl Kadie (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 461 (8), 43–52.

Chien, Yung-Hsin and Edward I. George (1999), "A Bayesian Model for Collaborative Filtering," working paper, Department of Information Systems, University of Texas, Austin.

Chung, Jaihak and Vithala R. Rao (2003), "A General Choice Model for Bundles with Multiple Category Products: Application to Market Segmentation and Pricing of Bundles," *Journal of Marketing Research*, 40 (May), 115–30.

Chung, Tuck S., Roland T. Rust, and Michel Wedel (2009), "My Mobile Music: An Adaptive Personalization System for Digital Audio Players," *Marketing Science*, 28 (1), 52–68.

Frühwirth-Schnatter, Sylvia (2004), "Estimating Marginal Likelihoods for Mixture and Markov Switching Models Using Bridge Sampling Techniques," *The Econometrics Journal*, 7 (1), 143–67.

Gelfand, Alan E. and Adrian F.M. Smith (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 972–85.

Jacobs, Robert A., Fengchun Peng, and Martin A. Tanner (1997), "A Bayesian Approach to Model Selection in Hierarchical Mixture-of-Experts Architectures," *Neural Networks*, 10 (2), 231–41.

Jordan, Michael I. and Robert A. Jacobs (1994), "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Computation*, 6 (2), 181–214.

Koren, Yehuda (2009), "The BellKor Solution to the Netflix Grand Prize," (accessed January 6, 2011), [available at www.netflixprize.com].

Little, Roderick J.A. and Donald B. Rubin (1987), *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Meng, X.L. and W.H. Wong (1996), "Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 6, 831–60.

Peng, Fengchun, Robert A. Jacobs, and Martin A. Tanner (1996), "Bayesian Inference in Mixtures-of-Experts and Hierarchical Mixtures-of-Experts Models with an Application to Speech Recognition," *Journal of the American Statistical Association*, 91 (435), 953–60.

Schafer, Ben J., Joseph Konstan, and John Riedl (1999), "Recommender Systems in E-Commerce," *ACM Conference on Electronic Commerce* (EC-99). New York: Association for Computing Machinery.

Stangl, Dalene K. and Donald A . Berry (2000), *Meta-Analysis in Medicine and Health Policy*. New York: Marcel Dekker Inc.

Takács, Gábor, István Pilászy, Bottyán Németh, and Domonkos Tikk (2009), "Scalable Collaborative Filtering Approaches for Large Recommender Systems," *Journal of Machine Learning Research*, 10 (June), 623–56.

Urban, Glen L, Fareena Sultan, and William Qualls (2000), "Placing Trust at the Center of your Internet Strategy," *Sloan Management Review*, 42 (1), 39–48.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (August), 282–94.

Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (August), 355–65.