Predictive validity across product categories with similar attributes.

Abou Bakar
Richard Lee
Cam Rungie

March 4, 2013

If the conference organizers make the offer, this paper will go onto the Journal of Choice Modeling

Abstract

In the study of consumer behavior the emphasis can be on attributes appearing in multiple product categories. The general area of diet, ‘lite’, low fat and low cal is an example of such an attribute but so too is brand and pack size. A single brand can hold a strong position in several categories. A consumer choosing bulk packs or family size may do so consistently. Even in heterogeneous willingness-to-pay studies the disutility for cost may or may not be consistent across categories. In the different categories the attribute might be described similarly but not the same. Low cost is an attribute that will manifest differently in each category due to differences in reference prices. Conversely, for some attributes the descriptions might be identical. The same brand name might be printed on all packs using the same font, color, size etc. Nevertheless, at least initially, a sensible approach is to assume that regardless of how similar the attribute descriptions might be there will be differences across categories to the consumer, as perceptions, and thus differences in the impact on choice. Consequently, the validity and accuracy of research using discrete choice experiments (DCE) is improved if the similar attributes are analyzed using separate DCEs for each category, rather than a single and combined experiment where the choice set includes products from several categories. The comparison of the separate DCEs can quite usefully start with fixed coefficient models but the core emphasis should be on random coefficients as they capture heterogeneity. The research questions can then ask if the heterogeneity is common across the DCEs for the similar attributes. Where a consumer has high utility for a particular level of an attribute is this consistent across the categories? The general research propositions is likely to be sometimes yes and sometimes no. When the categories are alike, such as two supermarket food categories, consistency is to be expected. Conversely, with unalike categories, such as the utility for a luxury attribute in both the automobile and petrol categories, inconsistency can be expected. A consumer may pay extra for luxury items when purchasing the car but then buy the petrol from a discount, low service and basic retailer.

In the paper, new methods for simultaneously modeling heterogeneity across categories are developed and demonstrated in cases where the separate DCEs are completed by the same sample of consumers. The modeling of multiple DCEs, with the same respondents, as a general method is achieved using latent variables. In other applications the DCEs might cover different choice situations for the same product, different times in a longitudinal study or, as in this application, different product categories. Over the multiple DCEs the latent variables allow the variance covariance matrix for the random coefficients to be fuller, i.e. specified to have off diagonal elements, while being parsimonious, identified, interpretable and the operationalization of specific research propositions. Latent variable models allow the heterogeneity to vary in common across the DCEs. The examples used in the paper analyze food and detergent categories in Pakistan. Different models are fitted and evaluated for competing propositions on the structure of heterogeneity and the extent to which it is common across the DCEs. The specification, identification and interpretation of the models are discussed. The models are validated through predictive validity. The
evaluation includes using the choices from a category to predict other choices in the same category as well as the choices in the other categories.

The paper concludes with a discussion. The study of consumer behavior focuses on patterns of purchase and usage over populations of consumers, be they individual, households, customers etc. The first step in consumer behavior research often is to spot the pattern and then the second step is to seek explanation. The predictive validity of the patterns is considered to be particularly important and some of the most valuable findings to date are patterns that have been established as generalizations over many data sets covering populations, markets and product categories. In contrast with the study of choice, which focuses on deconstructing individual decisions and is a within subject analysis, the study of consumer behavior focuses on populations and is the equivalent between subject analysis. Random coefficient models have particular potential in the field because the coefficients are specified to be constant within each consumer and to vary between consumers. The heterogeneity in the models is the pattern over the population of consumers. The models are mixed. Multiple DCEs, completed by the same respondents, modeled using latent variables, is a new method that has the potential to make a particularly useful contribution to the study of consumer behavior. Over populations of consumers heterogeneity can be evaluated for multiple DCEs. The method models patterns of choice for similar attributes across product categories.

Introduction

McFadden’s (1974) random utility theory (RUT) fixed coefficient model contributes substantially to the study of consumer behavior. It shows the average impact of the attributes of the alternatives in a choice set on the probability of selection, given a choice is made. For revealed preference data it deconstructs market share and for stated preference data it gives an indication of the influences on market share. The extension by Train (2003; 2009) to include random coefficients shows how this impact varies over consumers, which is useful in analyzing behavioral segmentation, improving model accuracy, and analyzing switching and consistency in repeated choice. This later area, known as behavioral loyalty, is of particular interest in consumer behavior. Historically the analysis has focused on the repeated choice of a single attribute, such as purchases of the brands in a product category (Goodhardt, Ehrenberg et al. 1984) or viewing of television programs (Sabavala and Morrison 1977). We describe here methods for analyzing repeated choice under a broader range of conditions. We show how the RUT models can be extended to evaluate the impact on switching and consistency of two or more attributes simultaneously and where the same attributes appear in two or more product categories. Thus the impact, of the attributes of the products in the choice set, on repeated choice can be evaluated.

The discussion is limited to zero order choice processes. For each consumer the repeated choices are independent. The systematic component of utility, for alternatives, attributes and levels is constant for each consumer but varies over consumers through the specification of heterogeneity. Consequently, a data set with multiple consumers and multiple choices per consumer will have patterns of serial correlation that might appear to be not zero order processes but are only due to heterogeneity.

From RUT, for consumer \( n \), alternative \( i \) has systematic component of utility \( V_{i,n} \) where the probability of selecting alternative \( i \) from choice set \( C \) is

\[
P_{i,n} = \frac{\exp(V_{i,n})}{\sum_{j \in C} \exp(V_{j,n})}
\]

Let there be \( k \) covariates \( X_{i,1,n} \) to \( X_{i,k,n} \) with coefficients \( \eta_{1,n} \) to \( \eta_{k,n} \) where

\[
V_{i,n} = \eta_{1,n} X_{i,1,n} + \ldots + \eta_{k,n} X_{i,k,n}.
\]
Over the population of consumers the coefficients \( \eta_{1,n} \) to \( \eta_{k,n} \) are random variables with distributions. In the random coefficient model \( \eta_{1,n} \) to \( \eta_{k,n} \) generally are either independent or have correlations directly estimated from the data. In latent variable model the \( \eta_{1,n} \) to \( \eta_{k,n} \) are specified to be linear functions of \( m \) latent variables \( \xi_{1,n} \) to \( \xi_{m,n} \) that also have distributions (Elrod 1988; Elrod and Keane 1995; Walker 2001; Ben-Akiva, McFadden et al. 2002; Rungie, Coote et al. forthcoming). Structural choice modeling (SCM) is a general form of latent variable model for which many properties are known (Rungie 2011; Rungie, Coote et al. 2011).

Random coefficient and latent variable models provide a range of joint and conditional probabilities of use in evaluating predictive validity. Two approaches to this evaluation are suggested. Firstly models can be compared; generally a model with greater predictive ability is preferred. Secondly, for the example considered here, models are evaluated on their capacity to predict choices in one product category, detergents, based on estimated by the models. The random coefficient and latent variable models provide a range of joint and conditional probabilities of choices per consumer. Random coefficient and latent variable models estimate the joint probability for all the alternatives selected by each consumer. Let a randomly nominated consumer select an alternative from each of \( a \) choice sets \( C_1 \) to \( C_a \). Let the selections be \( c_1 \) to \( c_a \). Then models estimates the joint probability

\[
\Pr\{c_1,\ldots,c_a \mid C_1,\ldots,C_a\}
\]

Due to the coefficients and latent variables having distributions over the population of consumers, the joint probability is not the product of the choice probabilities estimated singly for each of the alternatives selected.

Eq 3 can be used to estimate a range of other useful joint or conditional probabilities. For example \( \Pr\{c_1,\ldots,c_{a-1} \mid C_1,\ldots,C_{a-1}\} \), the joint probability of selections from choice sets 1 to \( a-1 \), as with Eq 3, is estimated by the models. Through, simple manipulation, conditional probabilities are estimated. The chance of \( c_a \) being selected from set \( C_a \) given all the other choices by the same consumer; i.e. given \( c_1 \) to \( c_{a-1} \), is

\[
\Pr\{c_a \mid c_1,\ldots,c_{a-1} \& C_1,\ldots,C_a\} = \frac{\Pr\{c_1,\ldots,c_a \mid C_1,\ldots,C_a\} \Pr\{c_1,\ldots,c_{a-1} \mid C_1,\ldots,C_{a-1}\}}{\Pr\{c_1,\ldots,c_a \mid C_1,\ldots,C_a\}}
\]

The joint probability in Eq 3 and the conditional probability in Eq 4 lead to four useful measures of the predictive validity of a model. The measures are geometric means of the choice probabilities for every alternative actually selected in the data but estimated under different conditions. When comparing two models fitted to the same data the higher these measures the greater the predictive validity.

- **ACP (average):** The average choice probability estimated singly for each alternative selected.
- **LCP (likelihood):** The choice probabilities are estimated jointly, from Eq 3, for the alternatives selected by each consumer.
- **WCP (within data set):** The choice probabilities are estimated singly for each alternative selected but conditional on the other alternatives selected by the same consumer.
- **XCP (across data set):** Below we discuss how two data sets from the same consumers can be modeled. The XCP, like WCP is conditional, but it examines the link between the two data sets. With XCP the choice probabilities are estimated singly for each alternative selected in one data set but conditional on all the alternatives selected by the same consumer in the other data set.
Formula

The total number of choices recorded in the data set is $A = \sum_{n=1}^{N} a_n$. Choice models can estimate the choice probability singly for every alternative selected. Taking the geometric mean gives a baseline evaluation

Eq 5 \[ ACP = \left( \prod_{n=1}^{N} \prod_{i=1}^{a_n} \Pr(c_{i,n} \mid C_{1,n}) \right)^{1/A} \]

Random coefficient and latent variable models estimate the joint probability for all choices by each consumer, as in Eq 3. The likelihood function uses the same estimate, leading to $LCP = L^{1/A}$. This can also be written as:

Eq 6 \[ LCP = \left( \prod_{n=1}^{N} \Pr \{c_{1,n}, \ldots, c_{a_n,n} \mid C_{1,n}, \ldots, C_{a_n,n} \} \right)^{1/A} \]

However, another comparison is available by examining the probability of each choice conditional on the other choices by the same consumer. Over the data set this is

Eq 7 \[ WCP = \left( \prod_{n=1}^{N} \prod_{i=1}^{a_n} \Pr(c_{i,n} \mid c_{1,n}, \ldots, c_{i-1,n}, c_{i+1,n}, \ldots, c_{a_n,n} \& C_{1,n}, \ldots, C_{a_n,n}) \right)^{1/A} \]

SCM can be applied to the combination of two or more discrete choice data sets collected from the same consumers but often under different conditions, such as different product categories (Rungie, Coote et al. 2011). Latent variables provide the link between the two data sets. With this matching of data sets it must be assumed that relevant covariates are omitted; even some covariates in one data set can be omitted from the other. Thus, the partitions of the model dealing with the different data sets will have different scales. The interpretation of some parameters will be affected but not all. In particular, the estimates of choice probabilities are not affected by scale. They are simply the best estimate of the choice probability given the information available from the covariates and model. (Similarly comparisons of log likelihood values provide valid evaluation of competing models.) It is these choice probabilities that provide the useful new outcomes in evaluating consumer behavior over product categories and in evaluating aspects of predictive validity.

With two data sets the choices are partitioned. For consumer $n$ let there be $a_{1,n}$ choices from in data set 1 and $a_{2,n}$ in data set 2. The choices are $c_{1,1,n}$, $c_{1,2,n}$ etc from choice sets $C_{1,1,n}$, $C_{1,2,n}$ etc and $c_{2,1,n}$, $c_{2,2,n}$ etc from choice sets $C_{2,1,n}$, $C_{2,2,n}$ etc. For data set 2 containing a total of $A_2$ choices:

Eq 8 \[ XCP = \left( \prod_{n=1}^{N} \prod_{i=1}^{a_{2,n}} \Pr(c_{2,i,n} \mid c_{1,1,n}, \ldots, c_{1,a_{2,n}}, \& C_{2,1,n}, \ldots, C_{2,a_{2,n}}) \right)^{1/A_1} \]

In the calculation of all four measures, $ACP$, $LCP$, $WCP$ and $XCP$, the computational problems of very small numbers can be avoided by working with natural logs. For example:

Eq 9 \[ LCP = \exp(LL/A). \]
Interpretation

An evaluation of competing models based solely on fit as assessed using log likelihood values, and associated AIC and BIC, is likely to be incomplete and ineffective. A model that only achieves better fit can have questionable contribution and can be difficult to interpret. However, there are exceptions. In some situations the fit may be a test of a hypothesis, such as in a likelihood ratio test between constrained and unconstrained models, where the hypothesis has considerable research importance. Nevertheless, good science is good prediction. Does the model predict better than other models? If so, what is the relevance? The measures, ACP, LCP, WCP and XCP help evaluate this predictive validity for competing models.

The measures can be compared, as a lower limit, to a naive model with uniform choice probabilities calculated from the choice set size. If the set has two alternatives then the naïve choice probability is 0.5. Similarly, set size of three has naïve choice probability of 0.33. Unfortunately, there is no easily established upper limit. Any one data set will almost never record every relevant covariate, and perfect prediction (if ever at all possible) will not be achieved. But of greater relevance is the construction of the choice sets. In revealed preference data the range of choice sets is limited. In stated preference the choice sets are purposefully designed to emphasize tradeoffs; i.e. choice probabilities closer to uniform. Studies are not designed to maximize the choice probabilities and if they were the accuracy of other estimates would be compromised. In most data sets the choice sets and covariates will inherently place an unknown upper limit on the choice probabilities. When two models are fitted to the same data the model with the greater choice probabilities is better but the same conclusion cannot be drawn when the models are applied to different data sets. The measures of choice probabilities are useful in evaluating predictive validity but have their limitations.

Two conclusions can also be drawn regarding model type. (i) LCP, WCP and XCP capture heterogeneity, but if the model does not, as is the case for the traditional fixed coefficient RUT model, then all three will be equal to the ACP. (ii) If a model fitted to two data sets from the same consumers does not specify linking between the data sets, as is the case for the traditional random coefficient model, then the XCP will be low. The model captures the heterogeneity in each data set but the two heterogeneities are not associated. As a result, there is no capacity to estimate the choices in one data set based on the choices in the other data set. This is demonstrated below.

The Data

Three choice experiments were conducted on the same sample of 241 consumers in Pakistan. The aim was to explore the consistency in preference for types of packaging across three product categories; soup, chocolate and detergent. Each experiment had two attributes, Color and Symbol each with two levels. Color had levels of Green, similar to the Pakistan national flag, and Red. Symbol specified if a graphic, the crescent moon, was Present or Absent on the pack. On-the-one-hand, the graphic is on the national flag and by attribution might influence choice in all three product categories more-or-less equally. On-the-other-hand the graphic is of particular relevance to the Islamic faith and, as the Halal guidelines of the faith focus more on food than non-food, may influence preference disproportionately. A balanced complete design was applied with choice set size of two. Each consumer completed six choice tasks for each of three product categories with a total of eighteen choices. Models of the data were designed to evaluate if
preference for Color and Symbol was consistent across the product categories. The three data sets were matched following the SCM methods of Rungie, Coote and Louviere (2011) discussed above. The covariates are specified in Table 1. Three models were fitted; the traditional Fixed Coefficient (6 parameters), the traditional Random Coefficient (12 parameters) and a model with two correlated Latent Variables (13 parameters) that linked the three product categories (the specification is in Figure 1).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Product Category</th>
<th>Attribute</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>Soup</td>
<td>Color</td>
<td>η₁</td>
</tr>
<tr>
<td>X₂</td>
<td>Soup</td>
<td>Symbol</td>
<td>η₂</td>
</tr>
<tr>
<td>X₃</td>
<td>Chocolate</td>
<td>Color</td>
<td>η₃</td>
</tr>
<tr>
<td>X₄</td>
<td>Chocolate</td>
<td>Symbol</td>
<td>η₄</td>
</tr>
<tr>
<td>X₅</td>
<td>Detergent</td>
<td>Color</td>
<td>η₅</td>
</tr>
<tr>
<td>X₆</td>
<td>Detergent</td>
<td>Symbol</td>
<td>η₆</td>
</tr>
</tbody>
</table>

There are six random coefficients, each a function of one of two latent variables

\[ V_{i,n} = η_{1,n}X_{i,1,n} + η_{2,n}X_{i,2,n} + η_{3,n}X_{i,3,n} + η_{4,n}X_{i,4,n} + η_{5,n}X_{i,5,n} + η_{6,n}X_{i,6,n} \]

Soup Colour \[ η_{1,n} = -2.20 + 3.73ζ_{n}^{\text{Color}} \]
Soup Symbol  \[ η_{2,n} = -0.85 + 3.24ζ_{n}^{\text{Symbol}} \]
Chocolate Colour \[ η_{3,n} = -1.80 + 3.47ζ_{n}^{\text{Color}} \]
Chocolate Symbol  \[ η_{4,n} = -0.77 + 2.47ζ_{n}^{\text{Symbol}} \]
Detergent Colour \[ η_{5,n} = -1.93 + 2.24ζ_{n}^{\text{Color}} \]
Detergent Symbol  \[ η_{6,n} = -0.90 + 2.64ζ_{n}^{\text{Symbol}} \]

\[ ζ_{n}^{\text{Color}} \text{ and } ζ_{n}^{\text{Symbol}} \] are latent variables (over the population of consumers n) with standard Gaussian distributions (mean fixed to 0 and variance fixed to 1) and correlation 0.56.

The model was fully identified as evaluated through (i) the usual practice of standardizing the latent variables, (ii) the Hessian and Information matrices having the required properties (invertible and leading diagonal all positive), and (iii) and the standard errors were all small.

We leave for documentation elsewhere the presentation of the full results, interpretation and discussion regarding the substantive research question on Islamic symbols. Here we discuss only the predictive validity, see Table 2. We examine the ability to predict the alternatives actually selected in the detergent product category.
As the choice set size was two, the measures in the table can be compared to the naïve baseline of 0.5. Random Coefficient and Latent Variable both have higher predictive validity than the naïve and Fixed Coefficient models.

As expected Random Coefficient has low XCP. It is not effective in predicting detergent choices based on soup and chocolate choices. The heterogeneity captured by the random coefficients in detergents is not associated with the heterogeneity in soup and chocolate. By comparison Random Coefficient has higher WCP. It can predict detergent choices based on other detergent choices by the same consumer.

Latent Variable has high WCP and XCP. When predicting detergent choice, the model is as effective in using soup and chocolate choices as it is in using other detergent choices. While we have left the discussion of the research question to be documented elsewhere the result supports the conclusion that utilities vary between consumers, but each consumer tends to be consistent across the three categories.

Latent Variable has greater predictive validity for comparisons between product categories. Whether or not higher result might be achievable with other data and other models is a matter of further research. Whether or not the result of 0.6 is good is a matter of judgment. We suggest, given that nature of the data with only two attributes and a balanced complete design, that 0.6 is a reasonable achievement and the model is creating predictive validity of relevance.

Table 2 provides the empirical evidence for two approaches to evaluating predictive validity. First it compares model and shows that Random and Latent Variable both perform well with the latter being better. Secondly it compares product categories and shows that for Latent Variable the choices in one category can be predicted from the choices in another. This gives empirical evidence supporting the argument that Latent Variable has greater predictive validity than the other models. An assessment of overall validity of a model should take into consideration additional issues and in particular (i) the theoretical support for the structure of heterogeneity, i.e. as specified in Figure 1, and (ii) the judgment as to whether or not the mean probabilities in the table, the XCP and WCP, are high enough. Nevertheless, the results in the table add to the validity of the Latent Variable model.

We note that in this example the measures are capturing within sample predictive validity for repeated choice. The approach however can be easily extended to holdout samples, based on consumers, where the parameter estimates from one sample of consumers, are used to estimate the choices by the holdout sample. Similarly, although computationally exhausting and of little
extra contribution, a boot strapping approach could be used where each consumer is held out, one at a time, with his/her choices predicted using parameter estimates from the N-1 remaining consumers in the sample. For a sample of 251 consumers, the ACP, LCP, WCP and XCP results will be rather similar but would then also have standard errors.

**Consumer Behavior**

Consumer behavior is the study of patterns of choice across consumers and in particular searches for empirical results that are generalizable over markets, populations and environments. It can be contrasted with areas such as psychology and the study of choice which focus more on the processes by which individual make decisions. The latter can be described as a within and the former as a between consumer analysis. Random coefficient and latent variable models add a between component to the within logit kernel of RUT. Consumer behavior has an interest in the distributions, known as mixing distributions, and structure of the random coefficients and latent variables. In particular generalizations are developed on the switching and consistency, known as behavioral loyalty, to be found in repeated choice. Traditionally behavioral loyalty is operationalised quite differently but is exactly equivalent to the conditional probabilities discussed above.

The consumer behavior literature contains results, discussions and empirical generalizations on the patterns for consumers repeated choices in revealed preference data and in particular on single attributes such as the repeat purchase of the brands in a category or the viewing of television programs. Measures have been developed for switching and consistency, referred to as behavioral loyalty, including the Index Polarization \( \varphi_i \), (Sabavala and Morrison 1977), the Dirichlet \( S \) (Goodhardt, Ehrenberg et al. 1984; Ehrenberg 1995), \( \theta \) (Bass 1969) and \( h \) (Kalwani and Morrison 1977). They are all mathematically linked, where \( \varphi_i = 1/(1+S) \), \( 0 < \varphi_i < 1 \) and the greater \( \varphi_i \) the greater the consistency in repeated choices. Empirical generalizations in the values \( \varphi_i \) and \( S \) have been developed (Driesener 2005).

The Index of Polarization, \( \varphi_i \), is a restatement of choice probabilities in Eq 3. Consider the case when consumers make two selections from the same choice set \( C \), with replacement and independence, then the probability of alternative \( i \) being selected on the first occasion is

Eq 10 \[ \text{Pr}\{c_1 = i \mid C\} = \pi_i \]

and on the second occasions conditional on the first is

Eq 11 \[ \text{Pr}\{c_2 = i \mid c_1 = i \& C\} = \rho_i \]

Fader and Schmittlein (1993) show that \( \rho_i = \pi_i + \varphi_i - \pi_i \varphi_i \). Solving gives an estimate of \( \varphi_i \).

Consequently, using appropriate choice models and Eq 10 and 11, behavioral loyalty can be calculated for any choice set, based on both either revealed preference (RP) or stated preference (SP) data provided multiple choices are recorded for each consumer. As shown above, different choice models can create different estimates of joint and conditional probabilities and so the models should be evaluated for predictive validity of repeated choice. It is to be expected that behavioral loyalty from comparable RP and SP studies will not be equivalent but generalizations may be identified on the differences. As it has been traditionally presented, the study of behavioral loyalty has analyzed a single attribute and has been limited in its ability to evaluate the choice set, separate several attributes or link similar attributes across product categories. However, as is shown above with latent variable RUT models these limitations are avoided. Behavioral loyalty is a property of the choice set and is just conditional choice. In as much as consumer behavior has an interest in the variations between consumers and in behavioral loyalty, choice models have a new, additional and continuing contribution to make.
Latent Variable Models

Over the population of consumers, the coefficients $\eta_{1,n}$ to $\eta_{k,n}$ in Eq 2 are random variables with covariance matrix $\Sigma$. The quality of the analysis of consistency and switching in repeated choice will be dependent on the quality, validity and parsimony of the specification of the structure of $\Sigma$. This is the contribution of latent variable models. By specifying and linking latent variables, prior knowledge on the structure of heterogeneity can be operationalized and tested while the number of parameters can be kept to a manageable and identifiable level. Predictive validity is not in the the foundation literature on latent variable models and there is even the view that, due to observations being endogenous and sometimes continuous, it might not be possible (Walker 2001). However, as seen above, with the newer forms of the models, and in particular SCM, this is no longer a limitation. It is now possible to evaluate the predictive validity of latent variable models, even when applied to two or more data sets from the same consumers. As a result the study of switching, consistency and behavioral loyalty is possible in a much wider domain than is traditionally found in the consumer behavior literature.

Conclusion

Four measures of predictive validity for models of repeated choice have been specified. Interpretation has been discussed in general and specifically to an empirical example where three data sets, being three product categories, are collected from the same consumers. In the example there is consistency in preference for two attributes, color and symbol, over the product categories, soup, chocolate and detergent. The paper shows that latent variables models for discrete choice can have predictive validity across multiple product categories and data sets. The outcome is an evaluation of heterogeneity, consistency in repeated choice, and behavioral loyalty for those attributes that are common across the multiple product categories.

References


