APPENDIX

Case Study: A Comparison of Models

Preliminaries
As part of our discussion, we have put together a simple comparison of Latent Class choice models to other forms of choice models. The comparison is not to do a definitive test of which model is best, but rather to demonstrate the power of modeling heterogeneity among respondents in explaining choice behaviors. We can understand choice behaviors much better using very simple models by capturing respondent heterogeneity. Improvement in computer and estimation technologies makes this process easier than ever before.

We compare three forms of models: the simple, aggregate, multinomial logit model, the latent class form of the same MNL model, and the hierarchical Bayes form of the same MNL model. The aggregate MNL model is developed in its most naïve form for comparative purposes only. By naïve form, we mean that the utility function is comprised of only generic attribute characteristics. No individual characteristics, nor complex model forms (such as the nested MNL, mixed logit, or GEV form are examined). The latent class models we develop use exactly the same utility function, but, of course, latent classes are uncovered that have clear parameter differences among the classes. Finally the HB MNL model fits parameters at the individual level; segments of one, so to speak.

The case is disguised. The business problem was that of an well-known hiking boot material manufacturer wishing to determine whether they could enter the well-established hiking boot market with their own branded hiking boot. One of their concerns was whether their entry into the market would simply cannibalize their existing share of the materials market by stealing market share away from the existing brands that use their material in their own boots. If the boot material manufacturer, Brand X, enters the market, from whom would they steal market share? Indeed, if they enter the market into which channels should they concentrate sales.

Channel and brand differentiate this market. Certain brands concentrate their sales in lower priced markets using lower priced channels. Generally, these channels offer lower priced boots catering to the more casual day hiker and outdoors enthusiast. Other brands concentrate their sales on the serious hiker, rock climber, and outdoors person. Higher priced boots are found in specialized sporting goods, outdoors stores, or even specialty shops. Brand X must determine which channel(s) they would enter if they decide to launch their own boot. We narrow our focus of analysis of respondent heterogeneity on the patterns of choices made among brand and channel.

In addition to these questions, brand X was also interested in the pricing of the boots, and whether they should exclusively introduce and new material feature.

Design and sample information

A stated choice modeling exercise was commissioned to answer these questions. The design attributes and their levels are described in detail in Table 1. The key attributes are: brand, channel, two different special features, and price. Channel and price posed a price conditional design problem because price and channel are correlated with each other in the market. Specific channels could not have the highest range of prices tested shown with them. Discount stores, for example, do not offer the highest priced, almost custom made boots. Specialty stores, conversely, seldom offer the lowest priced boots we wished to test. An addition design consideration included the fact that each boot brand also produces boots using their own materials as well as offering boots using brand X's materials, so the design had to incorporate these levels that are confounded directly with the brand into the design. Finally, one brand offered their boots at only their own specialty store.

Table A1. Design Attributes

Levels	Brand	Store	Performance Feature 1	Performance Feature 2	Price*
1	Merrell	Discount Store	Merrell Level	Standard	$50
2	Vasque	Catalog	Vasque Level	Special Upgrades	$75
3	Asolo	Sporting Goods	Asolo Level		$100
4	Salomon	Department Store	Salomon Level		$125
5	Brand Alpha	Specialty Outdoors Store	Alpha Level		$175
6			Alpha New Level		$225

The final design consisted of 72 choice sets of 6 branded alternatives plus a "None of these" alternative. Brands and channels were rotated across the columns of the design in a random fashion. The design was blocked into 4 versions of 18 choice tasks each.

The final sample consisted of 573 respondents who participated in an Internet survey. This sample was broken up into three parts for analysis purposes. First, we took the total data set and randomly selected 20% (112 respondents) and put them into a holdout sample data set. The holdout sample was not used in the model estimation. The remaining 80% (461 respondents) of the sample were used for model estimation after each respondent had two choice sets randomly withdrawn to use for holdout sets. The estimation data set finally consisted of 461 respondents who responded to 16 choice tasks.

The holdout sets data set consisted of 922 choices made by respondents in the estimation data set. We compare the predictive accuracy of the models we fit against these holdout sets as a measure of internal validity. While this is commonly done in market research, there are significant problems with using the "fit" figures to select the best model (Elrod, 2000). We show these figures strictly for comparison purposes. The holdout sample respondents were used to compare the "external validity" or transferability of the estimated models parameters. More on this comparison is discussed later.

Model Results

MNL Model

The naïve multinomial logit model was fit with 13 parameters: a simple, generic, linear effect for price (price/100); brand specific effects coding for brand and channel specific effect coding for channel; effects coding for the special features 1 and 2; and a dummy constant for the None alternative (1=None; 0=otherwise). In all, there were 13 parameters. The parameters are listed in the column label simple MNL in table 2. The fit statistics can be also be found in table 2. The model took less than one minute to estimate on a 1.4 GHz Pentium 4 PC using Windows 2000 and 756 mb of high speed memory.

The major results are that price is the most important attribute. The relative impact on the utility is far above that of any other attribute. Looking at brand, we see that brand X has a brand value almost as strong as Merrell, the strongest brand. Channel is less important than brand. The sporting goods and department stores are the strongest channels, followed by Outdoors and discount stores, with the Internet trailing last. The additional feature 2 and additional feature 1 have the second strongest impact on utility. The most immediate conclusion about brand and channel is that Merrell offered in a sporting goods store would generate the largest market share holding all else constant. Brand X would be a close second in the same channel. Notice with this coding of the MNL model we did not isolate the interactions between brand and channel, so the impact of each on utility is independent of one another.

Table 2
Models and Parameters

	Naïve MNL	8 Class Latent Class Model								HB MNL Model		
Variable	**1 Class**	**Class1**	**Class2**	**Class3**	**Class4**	**Class5**	**Class6**	**Class7**	**Class8**	**Mean$^+$**	**Lower 95%**	**Upper 95%**
Discount	0.0128	0.062	-0.084	-0.271	0.188	-0.737 *	-0.338 *	0.119	-0.271	-0.078	-1.312	0.953
	(0.045)	(0.115)	(0.307)	(0.155)	(0.172)	(0.176)	(0.164)	(0.212)	(0.285)	(0.671)		
Internet	-0.136 *	-0.135	0.264	0.041	-0.251	-0.411 *	-0.035	-0.078	0.231	-0.0709	-0.06072	0.427
	(0.036)	(0.111)	(0.280)	(0.104)	(0.147)	(0.120)	(0.111)	(0.195)	(0.207)	(0.310)		
Department	0.0436	-0.157	-0.350	0.054	-0.090	0.442 *	0.305 *	-0.005	0.351	0.076	-0.5986	0.7841
	(0.033)	(0.106)	(0.287)	(0.091)	(0.118)	(0.098)	(0.111)	(0.139)	(0.184)	(0.426)		
Sporting Goods	0.0657	0.333 *	-0.656	-0.039	-0.098	0.184	0.212	-0.015	-0.126	0.04478	-0.8387	0.9312
	(0.046)	(0.163)	(0.676)	(0.141)	(0.163)	(0.121)	(0.130)	(0.213)	(0.228)	(0.525)		
Outdoors**	0.014	-0.102	0.826	0.215	0.250	0.522	-0.144	-0.021	-0.186	0.028	-0.6941	0.8224
										(0.456)		
Merrell	0.2791 *	0.396 *	0.422	-0.195	0.721 *	-0.340 *	0.474 *	0.063	2.600 *	0.3256	-1.0447	1.85295
	(0.033)	(0.111)	(0.227)	(0.112)	(0.119)	(0.107)	(0.117)	(0.151)	(0.184)	(0.901)		
Vasque	-0.3137 *	-0.264 *	-0.883 *	-0.429 *	-0.416 *	-0.095	-0.282 *	-0.228	-0.966 *	-0.306	-1.0721	0.4637
	(0.036)	(0.117)	(0.319)	(0.115)	(0.123)	(0.095)	(0.118)	(0.144)	(0.249)	(0.486)		
Asolo	0.0611	0.103	-0.119	0.385 *	-0.049	0.329 *	0.162	-0.121	-0.114	0.2999	-0.593019	1.2507
	(0.034)	(0.109)	(0.235)	(0.113)	(0.147)	(0.097)	(0.121)	(0.182)	(0.192)	(0.591)		
NorthFace	-0.2864 *	-0.420 *	0.128	-0.364 *	-0.715 *	0.746 *	-1.337 *	-0.123	-0.897 *	-0.650	-2.1656	1.1964
	(0.045)	(0.212)	(0.497)	(0.156)	(0.222)	(0.095)	(0.189)	(0.182)	(0.281)	(1.075)		
Brand X**	0.260	0.184	0.452	0.603	0.458	-0.640	0.983	0.409	-0.623	0.33047	-0.7753	1.4024
										(0.679)		
Perf Feat 1 a	0.0884 *	0.127	-0.504 *	1.179 *	0.516 *	-0.089	0.326 *	0.211	-0.108	0.336	-0.34969	1.019
	(0.034)	(0.093)	(0.246)	(0.236)	(0.122)	(0.099)	(0.154)	(0.213)	(0.176)	(0.427)		
Perf Feat 1 a	0.3812 *	-0.072	0.781 *	2.384 *	0.445 *	0.361 *	0.814 *	0.736 *	0.273	0.762899	-0.337455	1.9377
	(0.036)	(0.125)	(0.367)	(0.260)	(0.122)	(0.092)	(0.134)	(0.227)	(0.154)	(0.721)		
Perf Feat 1 a**	-0.470	-0.055	-0.277	-3.563	-0.960	-0.272	-1.140	-0.948	-0.165	-1.099	-2.651	0.2638
										(0.880)		
Perf Feat 2	0.3388 *	0.305 *	0.832 *	0.758 *	0.271 *	0.190 *	0.202 *	1.551 *	0.311 *	0.62855	-0.1762	1.5738
	(0.018)	(0.058)	(0.155)	(0.063)	(0.066)	(0.051)	(0.080)	(0.165)	(0.096)	(0.505)		
Price	-2.4128 *	-5.670 *	-7.371 *	-3.152 *	-6.597 *	-1.039 *	-1.384 *	-3.631 *	-1.537 *	-5.401	-9.7698	-0.8767
	(0.061)	(0.284)	(0.778)	(0.210)	(0.441)	(0.139)	(0.208)	(0.372)	(0.290)	(2.766)		
None	-1.3075 *	-3.714 *	-2.260 *	-0.860 *	-7.887 *	-0.923 *	-3.222 *	-5.262 *	-1.642 *	-3.5625	-9.0885	1.9471
	(0.071)	(0.261)	(0.644)	(0.293)	(0.631)	(0.211)	(0.460)	(0.475)	(0.403)	(0.329)		
Class size		0.236	0.217	0.139	0.113	0.104	0.081	0.072	0.039			
Log Likelihood	-9,990.50		-7,219.30							-4056.100		
-2(LL0-LL)	2,966.40 13 df		8,508.80	216 df								
BIC	20,060.70		15,763.40									
R-Square	0.09		0.41							0.65		
Rho-Square	0.13		0.37							0.650		
MAE	1.35		0.90							0.64		
MSE	0.68		0.44							0.290		
CLE	0.54		0.34							0.2		
Est. Time	< 10 secs	6 min. 6 seconds on a 1.4 GHz Pentium 4 PC								2 hours 10 min. on a 1.4 GHz Pentium 4		

Latent Class Models

We tested several different latent class models. One particular model included the use of covariates to assist us in predicting latent class membership. This model used the same exact 13 parameters as the MNL model to predict utilities within each latent class. We tested the demographics of Income, gender, age and education as covariates. We purposely used these simple demographics to keep the problem small and because they are measures clearly available for non-respondents.

In LC models, covariates are used to predict class membership. They directly affect the overall utility of an alternative by affecting the class membership probability. As such, however, they do not directly influence any single parameter.

The Latent Gold Choice program was used to test a range of segment solutions, 2 to 20, models on the estimation data set using the 13 parameters in the utility function and the 4 covariates. The program was set to run 100 random start values, each with 100 iterations. The EM algorithm maximum was set to 1000 and a maximum of 50 Newton-Raphson iterations were allowed for final model fit. The best fitting model according to the BIC statistic had 8 classes (BIC=15,763). The parameters and fit statistics can be found in table 2. The program took 6 minutes and 6 seconds running on same machine as the MNL model.

The model is a significant improvement over the naïve MNL model. The McFadden Rho-square jumped from 0.13 to 0.37. All of the fit statistics show a remarkable improvement over the MNL model. The parameters are generally larger than those found in the MNL model. This is likely due to scale differences between the class-level results derived in the LC model as compared to the MNL. Tests for scale differences were not conducted.

The most interesting results begin to emerge when we examine the patterns of respondent heterogeneity across the latent classes. Table 3 depicts the proportion of the sample who prefer specific combinations of brand and channel. This table was constructed by examining the latent class parameters for brand and channel. We determined the specific combination of brand and channel parameter with the highest utility (brand parameter + channel parameter) within each latent class. The specific combination represents one cell in the matrix. We then placed the class size (in %) into the cell and added together all cells with more than one segment. We split class 2 equally between brands X and Merrell because the Brand X and Merrell brand parameters are so close.

Table A3
Brand x Channel Class Preferences

	Merrell	Vasque	Asolo	NorthFace	Brand X*	Channel Preference
Discount store					7.2%	7.2%
Over the internet						
Department store	3.9%				8.1%	12.0%
Sporting goods store	23.6%					23.6%
Outdoors store*	22.3%			10.4%	24.5%	57.2%
Brand Preference	49.8%			10.4%	39.8%	100%

In the case of the naïve MNL model, this matrix would consist of a single cell. In the latent class model, seven cells are occupied indicating much more diversity in the brand x channel preferences. While the Merrell brand is still the dominate brand, brand X now captures more of the market than what would have been predicted with the simple MNL. We should note these figures do not represent actual market shares, rather they represent that portion of the market with specific brand by channel preferences without estimating the actual market shares. It is an examination of the parameters across respondents. The actual substantive conclusions would need to be calculated using the market simulator

There is also a good deal of price heterogeneity across the classes as well. Classes 1, 2 and 4 all have price parameters 1 unit less than the weighted mean of the price parameters. These are predominately Merrell and Brand X classes. Classes 3 and 7 are relatively close to the average parameter, while classes 5, 6 and 8 are clearly much less price sensitive than any of the other classes. These classes include the Northface/Outdoors store combination and the Merrell/Department store combination.

The analysis of the covariates reveals that the covariates selected do not assist in a meaningful manner our ability to predict latent class membership. Table 4 shows the covariate parameters and their associated standard errors. Very few are significantly different from zero, even less than the number we would expect at chance. As such, these selected covariates will not assist us in making better predictions to the holdout sample data set. Of course, in the actual study, many more covariates were examined, and more were significant predictors of class membership, but still fewer than we would have liked. This suggests the posterior probability of class membership is not a function of the respondents' characteristics.

Table A4
8 Class Latent Class Covariate Parameters

Covariates	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8
Intercept	1.7672 *	1.7308 *	-0.6307	0.7199	-0.7182	-0.0715	-0.6165	-2.181
	(0.667)	(0.666)	(2.171)	(0.707)	(2.225)	(1.604)	(1.634)	(2.655)
Income								
$100k +	-0.628 *	-0.633 *	0.376	0.025	0.098	-0.288	0.118	0.932
	(0.320)	(0.315)	(0.325)	(0.350)	(0.334)	(0.500)	(0.613)	(1.261)
$25k to < $35k	0.126	0.249	-0.971	0.540	0.128	0.612	-1.564 *	0.882
	(0.315)	(0.311)	(0.515)	(0.384)	(0.408)	(0.401)	(0.731)	(1.282)
$35k to < $50k	-0.173	-0.234	0.129	-0.010	-0.716	-0.053	-0.143	1.200
	(0.279)	(0.275)	(0.311)	(0.330)	(0.387)	(0.366)	(0.440)	(1.232)
$50k to < $75k	0.067	-0.001	-0.124	-0.260	0.011	-0.820	0.585	0.542
	(0.263)	(0.267)	(0.319)	(0.332)	(0.328)	(0.421)	(0.412)	(1.255)
$75k to < $100k	0.608	0.619	0.590	-0.294	0.480	0.548	1.005	-3.556
	(0.714)	(0.713)	(0.738)	(0.776)	(0.741)	(0.763)	(0.785)	(4.744)
Age								
18	2.359	0.876	2.171	1.002	2.334	-3.034	-3.729	-1.978
	(1.849)	(1.880)	(1.891)	(2.012)	(1.908)	(7.475)	(7.518)	(7.479)
25	-0.814	-0.319	-0.248	-0.708	0.151	-0.154	1.137	0.955
	(0.431)	(0.429)	(0.461)	(0.525)	(0.457)	(1.538)	(1.546)	(1.539)
35	-0.366	-0.266	-0.134	0.372	-0.966	0.125	1.182	0.054
	(0.430)	(0.435)	(0.451)	(0.491)	(0.521)	(1.536)	(1.541)	(1.573)
45	-0.336	-0.319	-0.211	0.727	-0.762	0.886	-0.166	0.181
	(0.451)	(0.459)	(0.489)	(0.501)	(0.542)	(1.535)	(1.610)	(1.598)
55	-0.176	-0.157	-0.921	-0.075	-0.365	0.929	0.560	0.205
	(0.435)	(0.438)	(0.510)	(0.545)	(0.491)	(1.544)	(1.636)	(1.582)
65	-0.666	0.185	-0.657	-1.317	-0.391	1.248	1.016	0.582
	(0.533)	(0.482)	(0.626)	(0.871)	(0.602)	(1.551)	(1.631)	(1.606)
Gender								
Female	0.131	0.079	-0.012	-0.090	-0.065	-0.104	0.212	-0.151
	(0.114)	(0.113)	(0.141)	(0.156)	(0.156)	(0.183)	(0.185)	(0.231)
Male	-0.131	-0.079	0.012	0.090	0.065	0.104	-0.212	0.151
	(0.114)	(0.113)	(0.141)	(0.156)	(0.156)	(0.183)	(0.185)	(0.231)
Education								
College graduate	-0.507	-0.759	1.488	-0.243	1.237	-0.082	-0.746	-0.387
	(0.566)	(0.553)	(2.143)	(0.581)	(2.200)	(0.602)	(0.628)	(1.908)
Graduate school	-0.112	0.101	1.340	-0.293	1.998	-1.392	-1.887 *	0.245
	(0.594)	(0.572)	(2.157)	(0.645)	(2.207)	(0.840)	(0.958)	(1.921)
High school or less	0.261	0.411	2.008	-0.354	-3.471	0.791	-0.483	0.839
	(1.039)	(1.026)	(2.339)	(1.142)	(6.197)	(1.071)	(1.264)	(2.171)
Some college	-0.327	-0.642	1.535	-0.656	0.836	-0.661	-0.481	0.397
	(0.556)	(0.546)	(2.140)	(0.585)	(2.202)	(0.609)	(0.622)	(1.881)
Some graduate school	0.602	-0.209	-3.611	-0.288	2.216	-0.269	0.220	1.340
	(0.962)	(0.974)	(5.669)	(1.095)	(2.339)	(1.076)	(1.050)	(2.053)
Technical school	0.084	1.099	-2.760	1.835	-2.815	1.613	3.376	-2.432
	(2.361)	(2.289)	(9.226)	(2.301)	(9.232)	(2.326)	(2.359)	(9.162)

Because the covariates did not seem to improve our membership predictability, we went back to the model and fit it without any covariates. Again, we tested the range of 2 – 20 classes, but also the 35 classes and 50 classes model. The best fitting model according to the BIC statistic was the 14 class model. The parameters and fit statistics for this model are presented in Table 5.

In addition, we repeated the brand by channel heterogeneity analysis as seen in Table 6. The pattern is similar to those we saw with the 8 class model, only now there are more cells occupied by classes than before. This demonstrates there is increased parameter heterogeneity when we extend the analysis to 14 classes without covariates.

We do not report the parameters of the 35 and 50 class models in this paper. They were run primarily to demonstrate the ability of the latent class algorithm to carry its analyses beyond the recommended fit statistics when, for whatever reason, one wishes to over fit their data. Some of the fit statistics are reported in the validation section that follows.

Table A5
14 Class Latent Class Model

Variable	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10	Class11	Class12	Class13	Class14
Discount	-0.038	-0.019	0.671	0.331	-0.145	0.140	0.283	-0.085	-1.142 *	-0.932 *	-0.030	0.053	-6.531	-1.181
	(0.332)	(0.148)	(2.385)	(0.180)	(0.183)	(0.261)	(0.350)	(0.263)	(0.275)	(0.438)	(0.232)	(0.319)	(10.907)	(0.646)
Internet	0.264	0.008	0.301	-0.476 *	-0.146	-0.073	0.394 *	0.005	-0.478 *	-0.116	0.205	0.167	0.879	-0.265
	(0.268)	(0.133)	(2.390)	(0.162)	(0.136)	(0.203)	(0.189)	(0.214)	(0.166)	(0.262)	(0.170)	(0.235)	(2.738)	(0.454)
Department	-0.401	-0.197	0.202	-0.152	0.093	0.028	-0.075	-0.084	0.911 *	0.483 *	0.187	0.275	1.350	0.757
	(0.300)	(0.135)	(2.384)	(0.154)	(0.116)	(0.141)	(0.194)	(0.198)	(0.155)	(0.183)	(0.151)	(0.209)	(2.734)	(0.458)
Sporting Goods	-0.509	0.321	-3.130	0.119	0.420 *	-0.076	-1.022 *	-0.185	0.158	0.207	-0.045	-0.221	2.169	-0.169
	(0.611)	(0.186)	(9.406)	(0.182)	(0.133)	(0.250)	(0.341)	(0.243)	(0.174)	(0.285)	(0.174)	(0.261)	(2.735)	(0.486)
Outdoors**	0.684	-0.113	1.956	0.179	-0.222	-0.019	0.419	0.349	0.551	0.357	-0.317	-0.274	2.134	0.858
Merrell	0.586 *	0.504 *	-0.167	0.922 *	0.187	0.080	-0.327	0.450 *	0.072	-0.671 *	0.703 *	2.870 *	-0.895	0.296
	(0.240)	(0.149)	(3.158)	(0.142)	(0.136)	(0.159)	(0.259)	(0.179)	(0.148)	(0.277)	(0.143)	(0.226)	(0.593)	(2.133)
Vasque	-1.183 *	-0.211	0.683	-0.548 *	0.027	-0.267	-0.916 *	-0.453 *	0.154	-0.549 *	-0.647 *	-1.103 *	-0.433	-5.592
	(0.384)	(0.156)	(3.216)	(0.149)	(0.120)	(0.163)	(0.256)	(0.192)	(0.135)	(0.210)	(0.193)	(0.300)	(0.577)	(8.355)
Asolo	-0.118	0.327 *	0.676	-0.341 *	0.182	-0.043	0.698 *	0.025	0.008	0.000	0.331 *	-0.149	-0.771	3.916
	(0.262)	(0.151)	(3.199)	(0.161)	(0.120)	(0.148)	(0.189)	(0.194)	(0.133)	(0.210)	(0.147)	(0.235)	(0.582)	(2.118)
NorthFace	0.246	-0.793 *	-1.860	-0.487 *	-0.857 *	-0.161	0.401	-0.867 *	-0.171	0.551 *	-1.558 *	-1.141 *	3.127	1.862
	(0.479)	(0.301)	(12.721)	(0.207)	(0.186)	(0.221)	(0.322)	(0.286)	(0.190)	(0.254)	(0.286)	(0.343)	(2.188)	(2.123)
Brand X**	0.468	0.173	0.668	0.454	0.461	0.391	0.144	0.845	-0.063	0.669	1.172	-0.477	-1.028	-0.483
Perf Feat 1 a	-0.617 *	0.092	0.541	0.523 *	0.248	0.300	0.932 *	0.510 *	-0.377 *	1.161 *	0.935 *	0.038	0.045	0.542
	(0.278)	(0.120)	(0.419)	(0.134)	(0.134)	(0.270)	(0.305)	(0.222)	(0.149)	(0.553)	(0.241)	(0.201)	(0.222)	(0.492)
Perf Feat 1 b	0.936 *	-0.148	-1.624	0.471 *	0.472 *	0.742 *	1.654 *	0.431 *	0.536 *	4.379 *	1.116 *	0.225	0.676 *	1.279 *
	(0.398)	(0.152)	(1.006)	(0.146)	(0.129)	(0.214)	(0.360)	(0.186)	(0.125)	(0.674)	(0.242)	(0.171)	(0.212)	(0.507)
Perf Feat 1 c**	-0.319	0.056	1.083	-0.995	-0.720	-1.042	-2.586	-0.941	-0.159	-5.540	-2.051	-0.263	-0.721	-1.820
Perf Feat 2	0.826 *	0.273 *	0.961 *	0.232 *	0.155 *	1.410 *	1.200 *	0.170	0.345 *	1.206 *	0.208 *	0.227 *	0.151	0.290
	(0.175)	(0.078)	(0.235)	(0.077)	(0.066)	(0.142)	(0.131)	(0.103)	(0.100)	(0.131)	(0.080)	(0.110)	(0.097)	(0.176)
Price	-6.806 *	-4.739 *	-13.820 *	-5.919 *	-1.280 *	-3.371 *	-4.644 *	-7.270 *	-1.737 *	-3.477 *	-0.809 *	-1.405 *	-1.221 *	-1.668 *
	(0.913)	(0.376)	(1.686)	(0.452)	(0.192)	(0.315)	(0.583)	(0.798)	(0.246)	(0.394)	(0.226)	(0.340)	(0.236)	(0.525)
None	-1.653 *	-2.898 *	-7.241	-5.832 *	-0.356	-5.342 *	-2.783 *	-14.177 *	-4.079 *	0.212	-3.879 *	-1.330 *	0.088	0.988
	(0.747)	(0.383)	(3.837)	(0.527)	(0.249)	(0.684)	(0.681)	(5.573)	(0.529)	(0.672)	(1.092)	(0.445)	(2.211)	(2.224)
Class size	0.1936	0.1363	0.0956	0.0889	0.0766	0.072	0.0665	0.0543	0.0506	0.046	0.043	0.0333	0.0302	0.0131
Log Likelihood		-6,980.30												
-2(LL0-LL)		8,986.80	195 df											
BIC		15,156.60												
R-Square		0.46												
Rho-Square		0.39												
MAE		0.83												
MSE		0.41												
CLE		0.30												
Est. Time	8 min. 35 seconds on a 1.4 GHz Pentium 4 PC													

Table A6
Brand x Channel Class Preferences

	Merrell	Vasque	Asolo	NorthFace	Brand X*	Channel Preference
Discount store	8.9%				7.2%	16.1%
Over the internet					4.3%	4.3%
Department store	3.3%	5.1%	0.7%		4.6%	13.7%
Sporting goods store	13.5%			1.5%	7.7%	22.7%
Outdoors store*	19.4%	3.2%	10.6%	1.5%	8.6%	43.3%
Brand Preference	45.1%	8.3%	11.3%	3.0%	32.4%	100%

Hierarchical Bayes MNL model

Lastly, we fit a HB MNL model using the same 13 estimates as the MNL and LC models. We did include the same covariates as the 8 class LC model. In HB models, the covariates are used to predict the individual estimates of the model, not class membership as in the LC model. The results resemble 13 regression models with the regression of the individual-level estimate on the covariates. As such, the HB model covariates can have dramatically different impacts on the individual estimates.

The HB MNL model was fit using 10,000 burn-in iterations using non-informative prior estimate information. After the initial burn-in, we ran another 10,000 iterations saving every 10th iteration. This recommended interval reduces the autoregressive nature of the saved mean estimates for each respondent. The model took 2 hours and 10 minutes to fit on the same PC used to fit the other models. The log-likelihood for this model is -4,056.1 (MNL=-9,990.5; 8 class LC=-7,219.3). Obviously, the HB model is an improvement over the MNL and LC models. Table 2 also shows the mean estimates across respondents and their standard deviations. Other fit statistics are also shown in the table.

The mean estimates of the HB model are similar to the weighted average of the LC model parameters, but again, larger than the MNL model. The same hypothesis of the difference being due to possible scale differences holds for the HB model.

Examining the HB estimates, we see that most have large standard deviations. In fact, for all estimates except price the 95% confidence interval ranges from positive to negative values. The price confidence interval is always negative. This suggests enough respondent heterogeneity across estimates to see sign reversals on some attributes, as we saw in the LC model results.

We now have individual-level brand and channel estimates so we can examine each respondent's estimates and assign them to a cell in the brand by channel table we have used previously. Table 7 shows the increased amount of respondent heterogeneity across these two sets of estimates. All but three cells have a non-zero proportion of the sample who prefer the specific combination of brand and channel. The Asolo brand preference stands out much more. Merrell and brand X are still of similar brand strength spread across similar channels. This increased degree of captured respondent heterogeneity is borne out in an examination of the fit statistics and validation statistics in the next section.

Table A7
Brand x Channel Class Preferences

	Merrell	Vasque	Asolo	NorthFace	Brand X*	Channel Preference
Discount store	8.7%	0.4%	6.7%	0.4%	10.2%	26.5%
Over the internet	1.5%		4.8%		2.2%	8.5%
Department store	8.0%	0.9%	2.6%	3.9%	4.3%	19.7%
Sporting goods store	10.4%	0.2%	4.8%	1.3%	7.4%	24.1%
Outdoors store*	3.3%		7.6%	6.1%	4.3%	21.3%
Brand Preference	31.9%	1.5%	26.5%	11.7%	28.4%	100%

The covariates did not significantly assist us in estimating individual attribute estimates. While a few terms appeared to be significant, the overwhelming majority were not. As such, the demographic covariates we selected to demonstrate do not help us in fitting these models or estimates. As such, they do not help us in fitting our models to the holdout sample.

Model Validation

While we discuss model validation, we note that this is a single study subject to a set of poorly fitting covariates with which to test model validation. We present these results simply to demonstrate the improvement in fit we can obtain by modeling respondent heterogeneity. We fully expect the HB model to outperform the LC model in this case, and both the LC and HB models to outperform the naïve MNL model. The choice of which method to use should be driven by the client and researcher needs at the end of the project. If the client desires segments, or classes, of respondents and can act upon those classes in a more tangible way than with individual level estimates, then we recommend you using the LC models. If, however, the project needs require the individual level heterogeneity found in HB models, then one might prefer to use HB models. In a recent paper, Andrews and Currim (2002) suggested a study may result in poor HB model performance when the parameters are poorly identified. Other evidence suggests the HB models perform much better when you have more data per respondent than less. LC models are less subject to these issues than HB models.

Model validation is a tricky issue. Terry Elrod in his Sawtooth, 2001 paper indicated that simple measures of holdout set validation traditionally used are inappropriate measures for determining the best fitting model. He suggests using holdout sample cross-validation procedures for determining the best fitting model. Comparisons of model fit to the holdout sets/tasks of each respondent in the estimation data set are not as good as comparisons to a completely independent holdout sample. In our example, we did not perform a 4-fold cross validation holdout sample test as recommended by Elrod. We did conduct a single holdout sample test.

In addition, he recommends replacing the traditional measures of mean absolute and mean squared error (MAE and MSE) and the classification "error" rates (CLE) with the calculated log-likelihood for the models based upon the holdout sets and holdout samples. We report the mean log likelihood, multiplied by -1 to make the value positive. Table 8 displays the formulas we used to estimate these measures.

Table A8
Measures of model fit used

Mean absolute and mean squared error

$\text{MAE} = \Sigma_n \text{ABS}(\text{prob}_{\text{actual}} - \text{prob}_{\text{pred}}) / n$

$\text{MSE} = \Sigma_n (\text{prob}_{\text{actual}} - \text{prob}_{\text{pred}})2 / n$

Classification error

CLE = 1 - Hit Rate;

Where:

Hit Rate = proportion of observations correctly predicted by the model.

Predicted choice is the alternative having the highest predicted probability of selection

Mean log-likelihood

-MLL = -1 * [LL(b) / n]

We have three data sets with which to compare fit measures: the estimation data set (N=461, 16 sets/respondent), the holdout sets/tasks (N=461, 2 sets/respondent randomly drawn from each respondent, and the holdout sample (N=113, 18 sets/respondent). Table 9 presents the results of our fit measures.

Table A9
Goodness of Fit Measures

	Estimation Data Set					Holdout Sets				Holdout Sample			
Model	**Time (in min.)**	**-MLL**	**MSE**	**MAE**	**CLE**	**-MLL**	**MSE**	**MAE**	**CLE**	**-MLL**	**MSE**	**MAE**	**CLE**
MNL w/o covariates	< 0.1	1.355	0.679	1.354	0.539	1.325	0.674	1.346	0.546	1.362	0.678	1.353	0.539
LC 8 class w/ covariates	6	0.857	0.440	0.895	0.335	0.959	0.483	0.932	0.370	1.415	0.709	1.363	0.568
LC 14 class w/o covariates	8.5	0.778	0.402	0.827	0.309	0.913	0.459	0.877	0.348	1.354	0.676	1.344	0.538
HB w/ covariates	130	0.550	0.291	0.641	0.204	0.861	0.435	0.788	0.318	1.744	0.796	1.253	0.579
LC 35 class w/o covariates	~17	0.665	0.350	0.710	0.256	-- Not examined --				-- Not examined --			
LC 50 class w/o covariates	~25	0.630	0.336	0.679	0.248	-- Not examined --				-- Not examined --			

The latent class models and the HB models clearly outperform the naïve MNL model in the estimation data set. The MAE and –MLL values drop quickly as we fit more parameters to this data. Estimation of the LC model is faster than the HB model, but the fit measures are clearly superior with the HB model. All three measures: the MAE, MSE, and CLE measures

In addition to the 8 class and 14 class models, we also fit a 35 and 50 class LC model. For these models, we had to rely on the use of only the EM algorithm and the use of Bayesian constants in order for the models to converge. We only report their estimation data set fit measures. There is discussion in the literature that HB models are over fitting the parameters to the data. That has you fit more and more parameters there ability to predict to other data sets may not be as good as their fit to the estimation data. We fit these extreme cases of the LC model to demonstrate that improved fit measures can be derived by over fitting, even with LC models. We do not attempt to test the models against the holdout tasks or holdout sample because we believe they are over fitted models.

In the holdout tasks data set, the HB model outperforms both LC models, but the level of its own performance has dropped considerably. The MSE and CLS measures dropped by over 30% and the MAE dropped by 18%. The fit measures for the LC models, however, are considerably more stable. The MSE and CLE measures dropped by approximately 12% each, while the MAE measure dropped by only 6%. This could be considered evidence of the degree of over fitting by both models.

The results of the holdout sample are very disappointing to us. Only the 14 class LC model did better than the naïve MNL model. Even its improvement is so small it is virtually indistinguishable form the naïve MNL model's fit. This is a result of having poor covariate predictors in all the models. Because none of the covariates helps us predict class membership, or HB model parameters, we have no means to utilize the increased knowledge of respondent heterogeneity in making predictions to an external data set. In fact, by not having any covariate predictability, the best prediction we can make for the external data set is the aggregate, or average, predictions made by the naïve MNL model.

One implication of this poor attempt at predicting to an external data set is, that an LC or HB model fit without significant covariates which can explain the respondent heterogeneity within the estimation data set, is restricted to making inferences to the sample from which we collected the data and the population it represents. Attempting to use that model to make inferences beyond the sample's domain, will be no better than the naive MNL model. Given our poor explanation of the respondent heterogeneity, we dare not attempt draw any conclusions about the external validity of these models.