Tutorial #5: Analyzing Ranking Data

Choice tutorials 1-4 all dealt with the analysis of *first* choices among sets of alternatives. In applications where information is also available on additional choices -- 2nd choice, 3rd choice, last choice, etc. -- improved efficiency of the part-worth utility estimates is possible by taking this additional information into account. In such cases, Latent GOLD Choice allows the utilization of the *sequential* logit model, a generalization of the conditional logit model, to account for the additional choice information.

The way the sequential logit model works is that the first choice is analyzed as usual based on the conditional logit model. If the model specified in Latent GOLD Choice is set to a 'Ranking' Model, after the first record within a set, any additional records are assumed to be associated with a 2nd choice, 3rd choice, etc. A 2nd choice is considered to be a first choice from the set of alternatives that *excludes* the 1st choice, and so on for the 3rd, 4th and additional choices. For ranking models, Latent GOLD Choice automatically excludes these prior choices from the consideration set of alternatives used for a current choice.

This tutorial deals with *full* ranking data obtained from a real bank segmentation study as described in Kamakura, Wedel, and Agrawal (1994), "Concommitant variable latent class models for conjoint analysis", *International Journal of Research in Marketing*,11, 451-464. The data was provided for our use by Wagner Kamakura.

This tutorial illustrates the use of the Latent GOLD Choice program to analyze ranking data.

You will:

- Identify 4 segments that differ in the importance placed upon various checking account attributes.
- Interpret output in the context of rank-order preference data.
- Use concomitant variables ("covariates") to predict and describe these segments.

In our *next* tutorial (tutorial #6), we will examine the performance of *partial* rank information with these data. Partial rankings, such as the first and last choices (as obtained in 'best-worst' or 'max-diff' designs) require less effort on the part of responders to complete than *full* ranking tasks.

Bank Segmentation Example

Respondents were asked to rank 9 checking account alternatives (from most to least preferred). These alternatives were created from the following attributes using a fractional factorial design:

MINBAL – The minimum balance required to waive a monthly service fee.

- $0
- $500
- $1000

COSTPCH – The amount charged per check issued during the month.

- $0.00
- $0.15
- $0.35

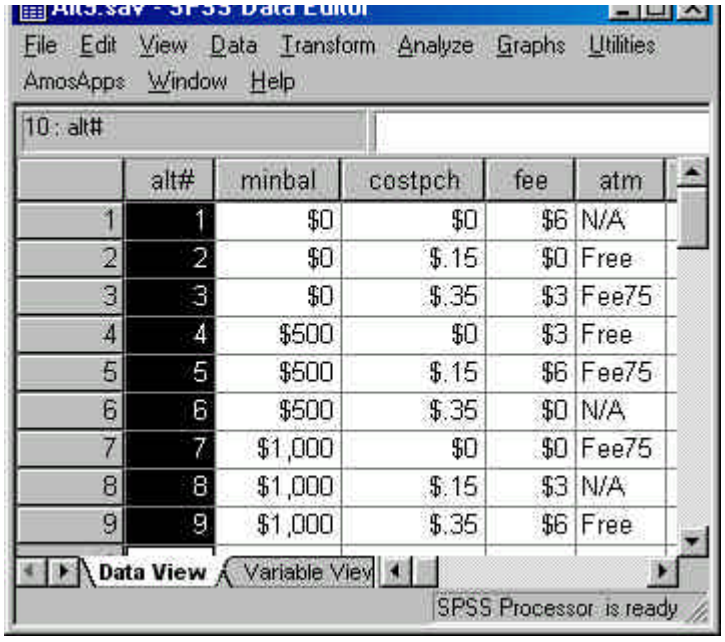
FEE – Monthly service fee.

- $0
- $3
- $6

ATM – Availability and cost for using automatic teller machines in a network of supermarkets.

- not available
- available and free
- available but costs $0.75 per transaction

The 9 alternatives are:



	alt#	minbal	costpch	fee	atm
1	1	$0	$0	$6	N/A
2	2	$0	$.15	$0	Free
3	3	$0	$.35	$3	Fee75
4	4	$500	$0	$3	Free
5	5	$500	$.15	$6	Fee75
6	6	$500	$.35	$0	N/A
7	7	$1,000	$0	$0	Fee75
8	8	$1,000	$.15	$3	N/A
9	9	$1,000	$.35	$6	Free

In addition to the ranking information, the following concomitant variables were available on respondents who had been customers of the bank for at least one year:

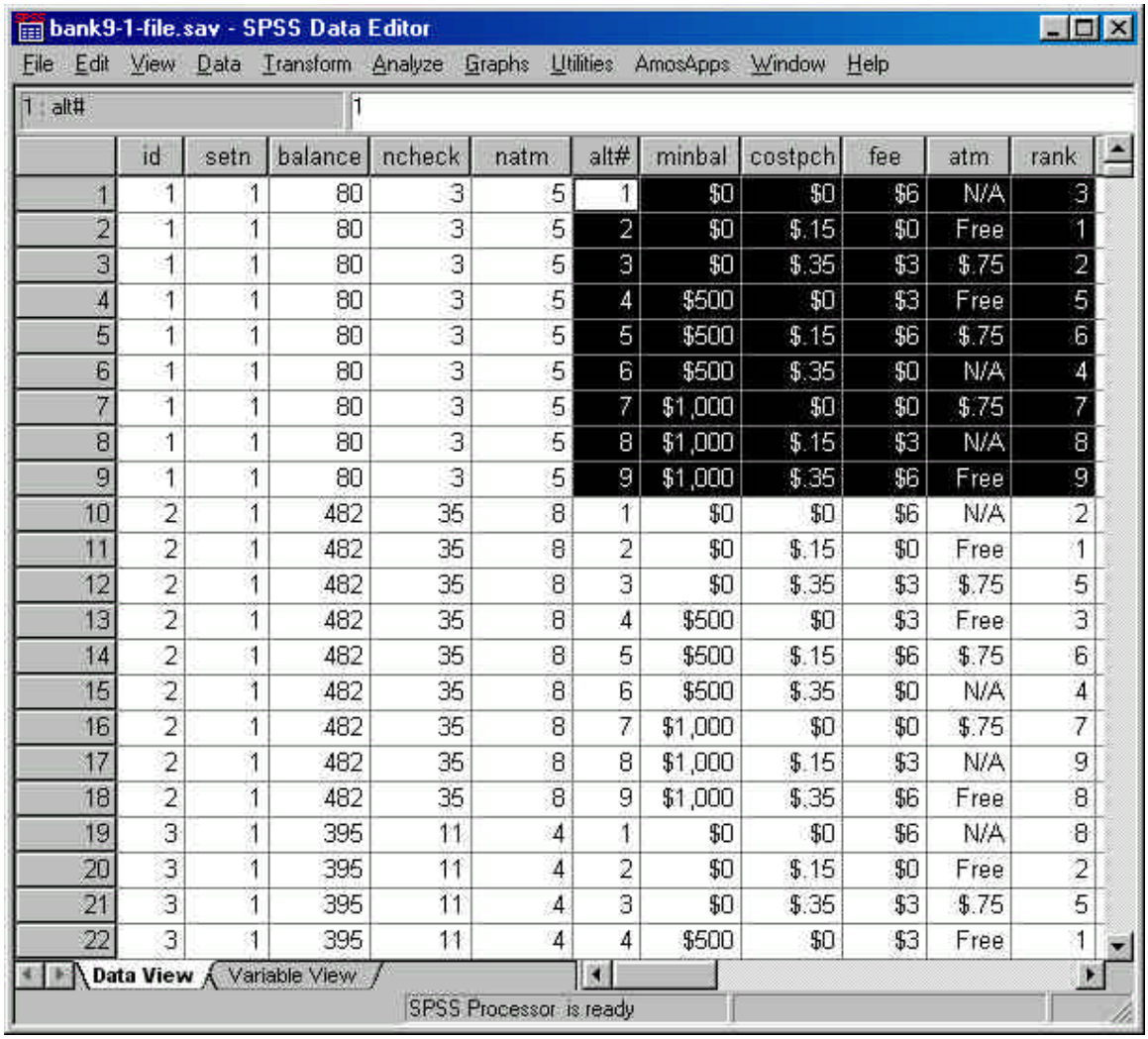
BALANCE – Average balance kept in the account during the past 6 months, earning 5.5% interest.

NCHECK – Number of checks issued per month in the past 6 months (at no charge).
NATM – Number of ATM transactions per month (all ATM machines).

These variables will be used as covariates in our model.

The sample consists of 256 bank customers for whom complete data were available on the covariates.

For this tutorial we will use 'bank9-1-file.sav', which illustrates the 1-file format. This file contains 9 records per respondent, corresponding to the 9 alternatives. The variables consist of a) a case ID, b) the covariates associated with the case, c) the levels of the 4 attributes associated with each alternative, and d) the rank order assigned to that alternative. The first 22 records on this file are shown below. The variable SETN (= 1 for all records) indicates that there is only a single choice set. (The next tutorial illustrates a ranking situation involving 2 choice sets.)



	id	setn	balance	ncheck	natm	alt#	minbal	costpch	fee	atm	rank
1	1	1	80	3	5	1	$0	$0	$6	N/A	3
2	1	1	80	3	5	2	$0	$.15	$0	Free	1
3	1	1	80	3	5	3	$0	$.35	$3	$.75	2
4	1	1	80	3	5	4	$500	$0	$3	Free	5
5	1	1	80	3	5	5	$500	$.15	$6	$.75	6
6	1	1	80	3	5	6	$500	$.35	$0	N/A	4
7	1	1	80	3	5	7	$1,000	$0	$0	$.75	7
8	1	1	80	3	5	8	$1,000	$.15	$3	N/A	8
9	1	1	80	3	5	9	$1,000	$.35	$6	Free	9
10	2	1	482	35	8	1	$0	$0	$6	N/A	2
11	2	1	482	35	8	2	$0	$.15	$0	Free	1
12	2	1	482	35	8	3	$0	$.35	$3	$.75	5
13	2	1	482	35	8	4	$500	$0	$3	Free	3
14	2	1	482	35	8	5	$500	$.15	$6	$.75	6
15	2	1	482	35	8	6	$500	$.35	$0	N/A	4
16	2	1	482	35	8	7	$1,000	$0	$0	$.75	7
17	2	1	482	35	8	8	$1,000	$.15	$3	N/A	9
18	2	1	482	35	8	9	$1,000	$.35	$6	Free	8
19	3	1	395	11	4	1	$0	$0	$6	N/A	8
20	3	1	395	11	4	2	$0	$.15	$0	Free	2
21	3	1	395	11	4	3	$0	$.35	$3	$.75	5
22	3	1	395	11	4	4	$500	$0	$3	Free	1

Notice that the 9 alternatives appear in sequential order (alt#1 – alt#9) for each case. In general, the ordering of the alternatives is not important so long as the attribute levels associated with each of these records is appropriate for that alternative. The variable named 'RANK' in the file represents the rank order of the alternative. For example, case #1 ranks alt#2 first (RANK = 1 for the second record which corresponds to alt#2).

Setting up the analysis

- To setup the 4-class model, from the main menu choose:

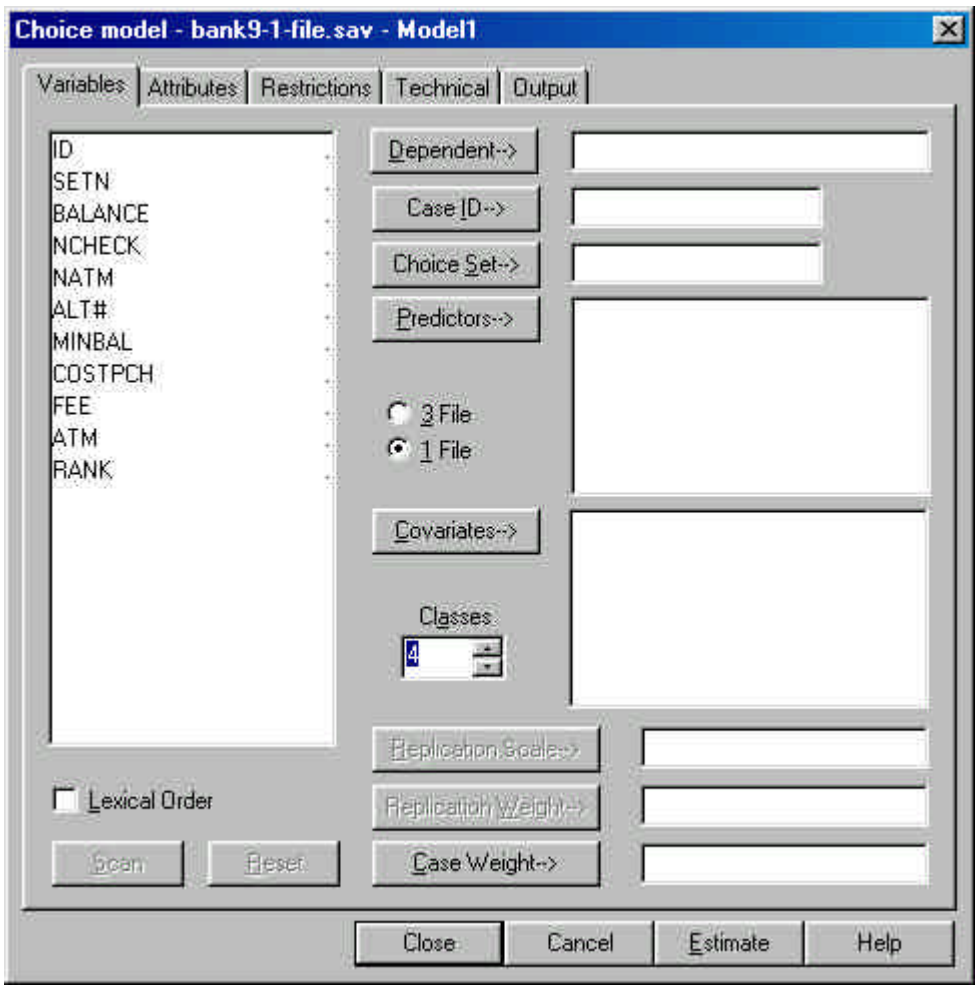
File
Open

- From the Files of type drop down list, select SPSS system files (.sav) if this is not already the default listing.

All files with the .sav extensions appear in the list.

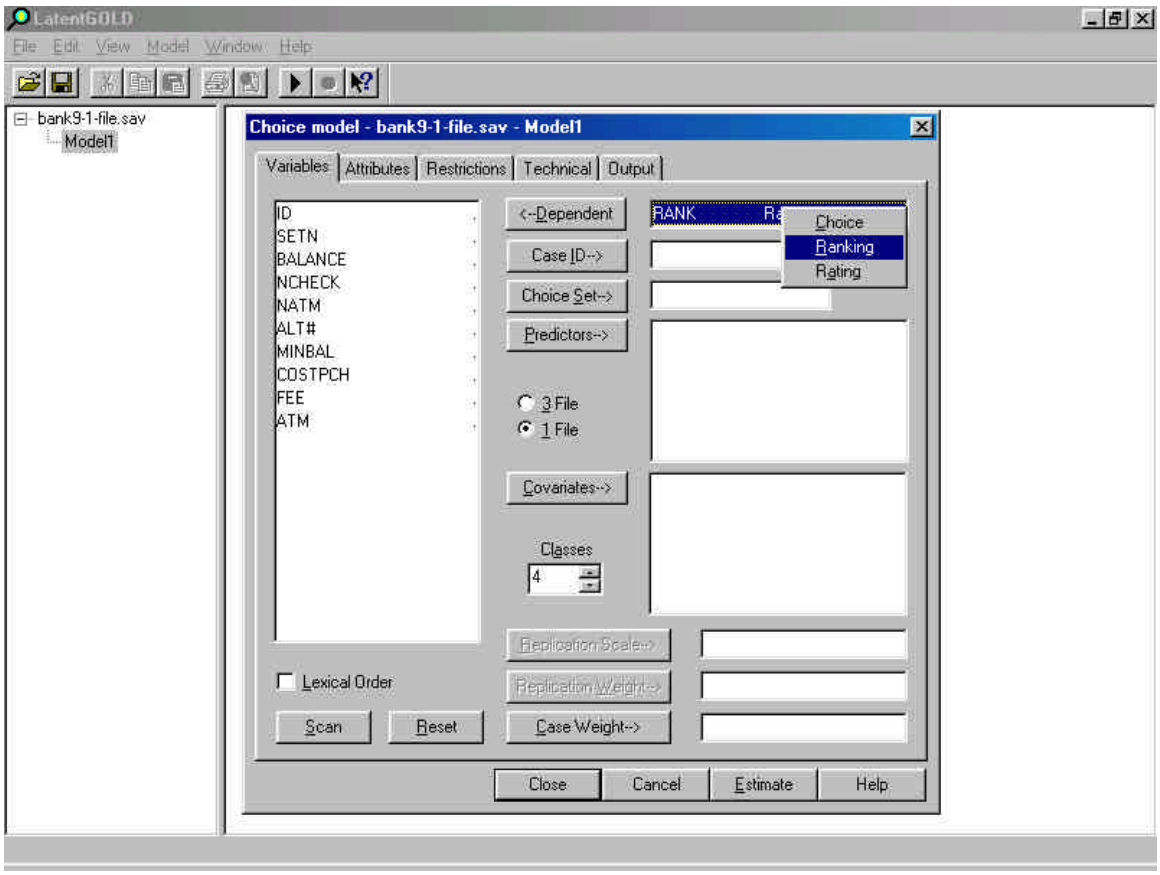
- Select 'bank9-1-file.sav' and click Open
- Click on 'Model1' and select Edit to open the model setup window
- Click the radial button to the left of '1 File' to indicate the 1-file format is being used
- Change the '1' to '4' in the Classes box to specify 4 classes

Your setup window now looks like this:



To assign the appropriate settings to the variables,

- Click on the variable named 'RANK' in the Variable list and click on the button marked 'Dependent' to move it to the Dependent Variable box
- Right click in the Dependent Variable box to retrieve the pop-up menu
- Select 'Ranking' to indicate that we will be estimating a Ranking model

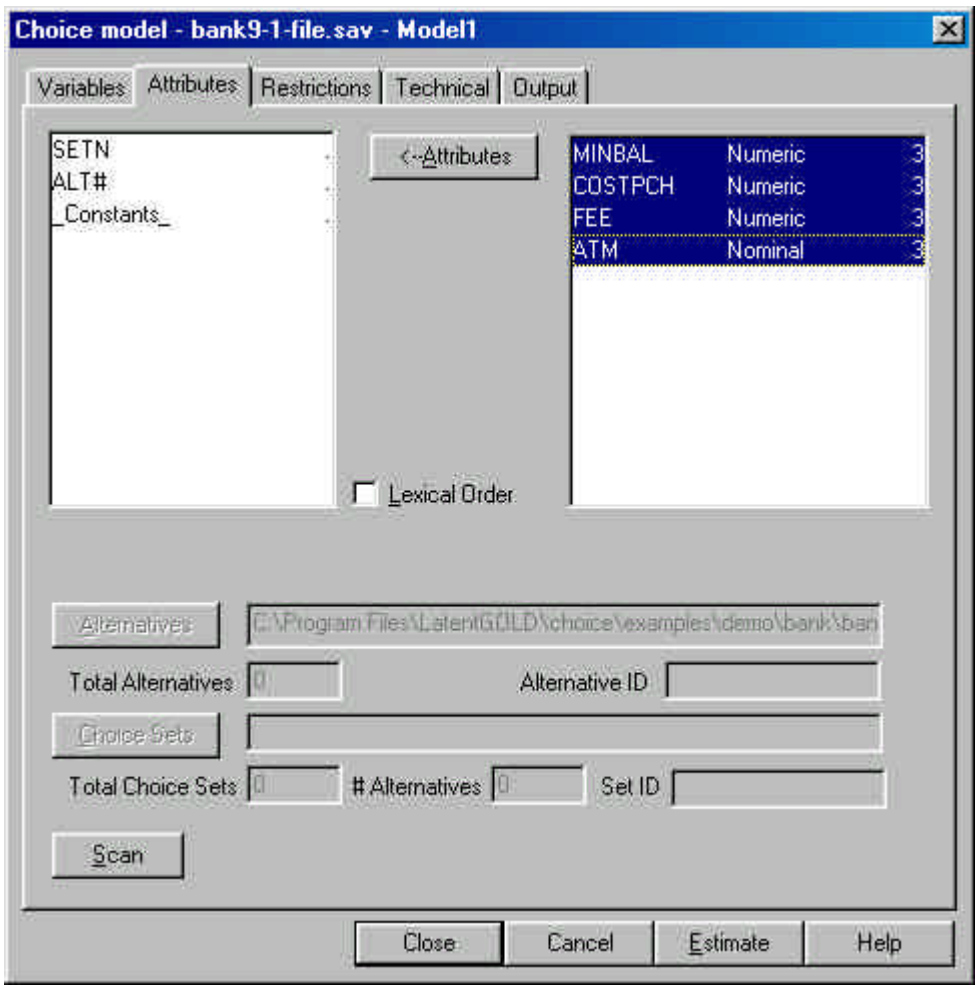


(Notice that for a 1-file format, the assignment of the choice set variable SETN to the Choice Set box is optional. By default, for a 1-file format the program assumes that there is only 1 set which is the case here.)

To complete the setup,

- Move ID to the Case ID box
- Move the variables BALANCE, NCHECK, and NATM to the Covariates box
- Now Click the Attributes tab to Open the Attributes Window
- Move the 4 attribute variables (MINBAL, COSTPCH, FEE and ATM) to the Attributes box
- Click Scan to scan the file

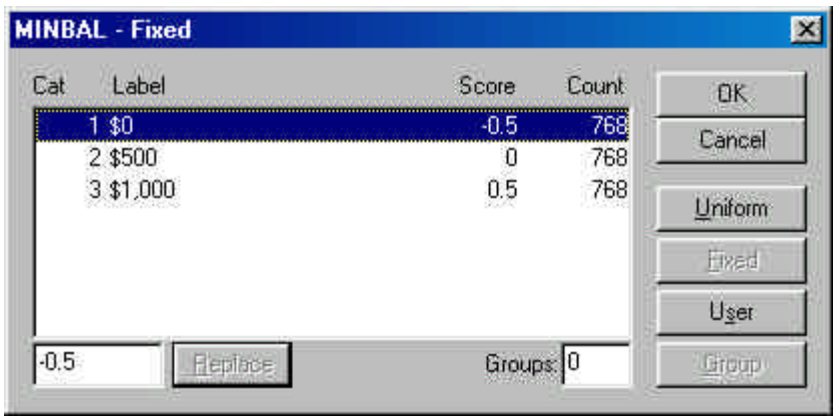
Your screen now looks like this:



Notice that by default the attribute ATM is treated as 'Nominal' while the others are treated as 'Numeric'. That is because in the SPSS .sav file, ATM is defined as a string (character) variable, while the others take on the equidistant numeric values used by Kamakura et. al.

To view the numeric values assigned to these attributes, you may open the Score box.

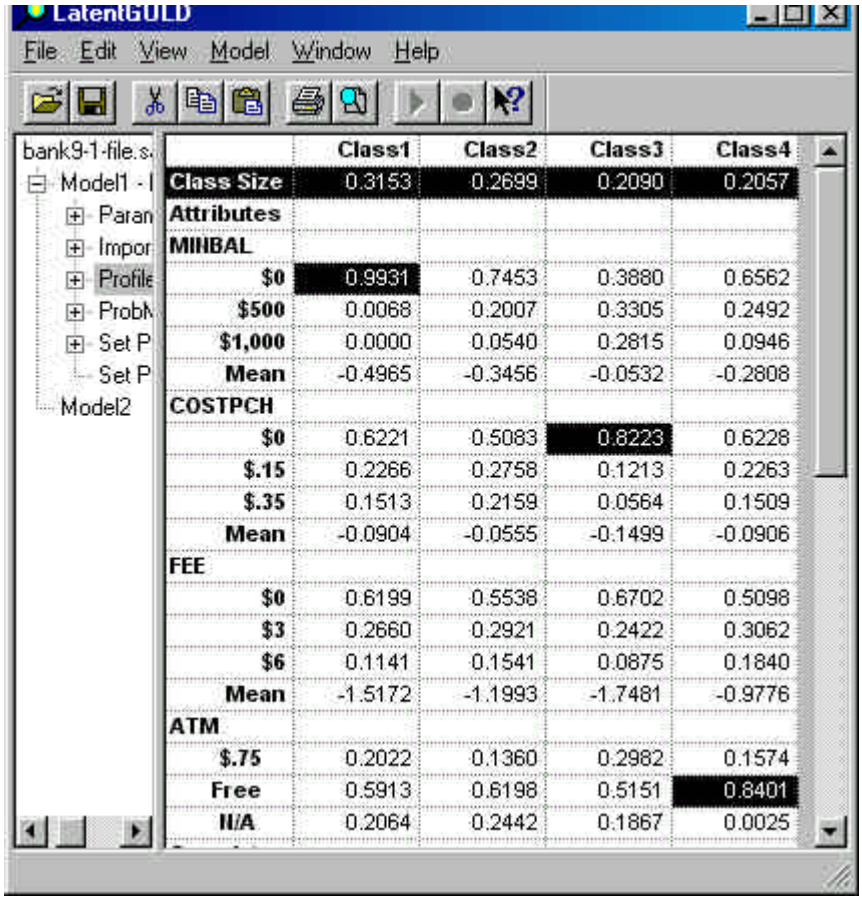
- Double click on MINBAL to open the score box for this numeric attribute



- Click Estimate to estimate the 4-class model

When the estimation has completed,

- Click Profile to view the Profile Output

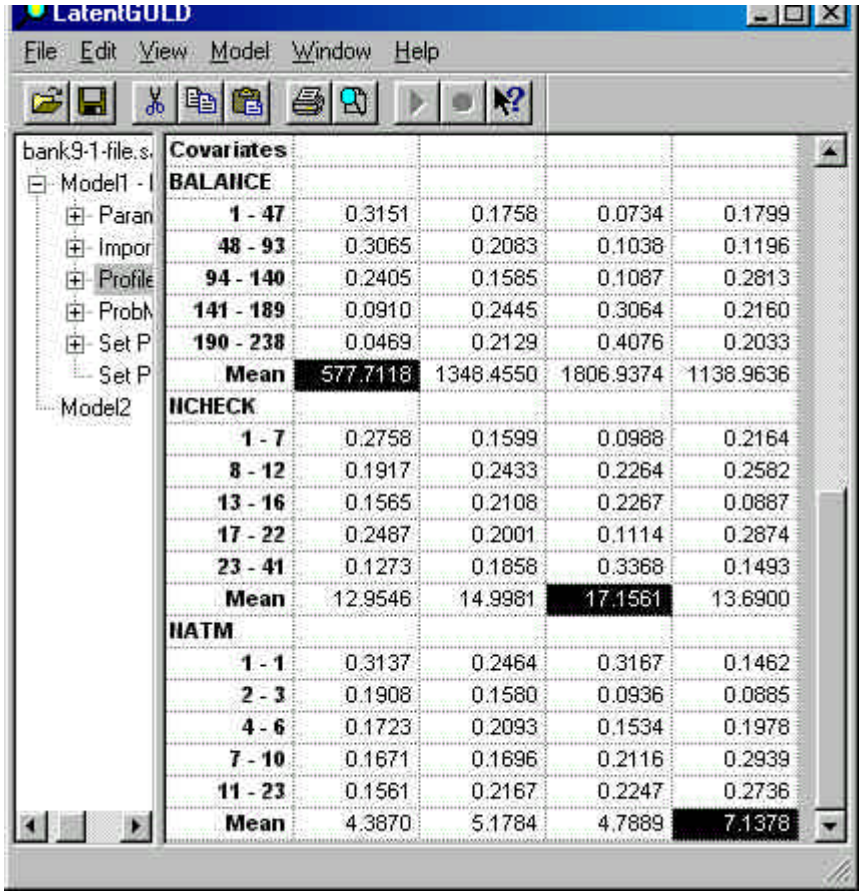


	Class1	Class2	Class3	Class4
Class Size	0.3153	0.2699	0.2090	0.2057
Attributes				
MINBAL				
$0	0.9931	0.7453	0.3880	0.6562
$500	0.0068	0.2007	0.3305	0.2492
$1,000	0.0000	0.0540	0.2815	0.0946
Mean	-0.4965	-0.3456	-0.0532	-0.2808
COSTPCH				
$0	0.6221	0.5083	0.8223	0.6228
$.15	0.2266	0.2758	0.1213	0.2263
$.35	0.1513	0.2159	0.0564	0.1509
Mean	-0.0904	-0.0555	-0.1499	-0.0906
FEE				
$0	0.6199	0.5538	0.6702	0.5098
$3	0.2660	0.2921	0.2422	0.3062
$6	0.1141	0.1541	0.0875	0.1840
Mean	-1.5172	-1.1993	-1.7481	-0.9776
ATM				
$.75	0.2022	0.1360	0.2982	0.1574
Free	0.5913	0.6198	0.5151	0.8401
N/A	0.2064	0.2442	0.1867	0.0025

Note the following segment preference differences exhibited by these data. Segment 1, accounting for about a third (.3153) of the customers, are more likely than the other

segments to prefer alternatives offering a 'No minimum balance' option (the re-scaled part-worth utility associated with MINBAL = $0 is .9931). Segment 3, accounting for about 21% of the customers) gives more weight to no check-writing cost (re-scaled parameter associated with COSTPCH = $0 is .8223). In contrast, Segment 4, accounting for 21% of the customers, is more influenced by free ATM availability (ATM = 'Free'). The preferences for Segment 2, accounting for the remaining 27% of customers, falls somewhere between these other segments in their preferences.

- Scroll down to display the covariate distributions

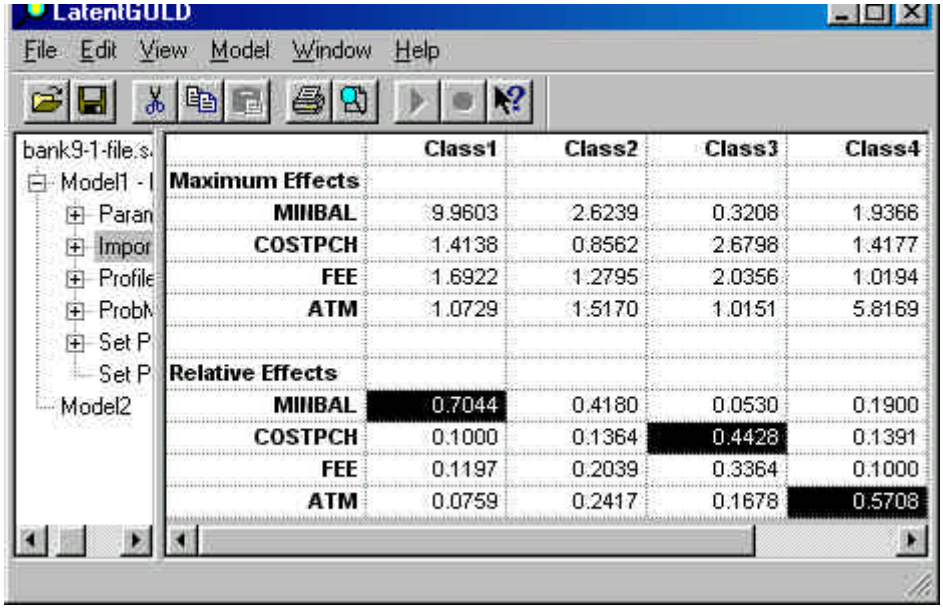


Covariates				
BALANCE				
1 - 47	0.3151	0.1758	0.0734	0.1799
48 - 93	0.3065	0.2083	0.1038	0.1196
94 - 140	0.2405	0.1585	0.1087	0.2813
141 - 189	0.0910	0.2445	0.3064	0.2160
190 - 238	0.0469	0.2129	0.4076	0.2033
Mean	577.7118	1348.4550	1806.9374	1138.9636
NCHECK				
1 - 7	0.2758	0.1599	0.0988	0.2164
8 - 12	0.1917	0.2433	0.2264	0.2582
13 - 16	0.1565	0.2108	0.2267	0.0887
17 - 22	0.2487	0.2001	0.1114	0.2874
23 - 41	0.1273	0.1858	0.3368	0.1493
Mean	12.9546	14.9981	17.1561	13.6900
NATM				
1 - 1	0.3137	0.2464	0.3167	0.1462
2 - 3	0.1908	0.1580	0.0936	0.0885
4 - 6	0.1723	0.2093	0.1534	0.1978
7 - 10	0.1671	0.1696	0.2116	0.2939
11 - 23	0.1561	0.2167	0.2247	0.2736
Mean	4.3870	5.1784	4.7889	7.1378

Latent GOLD Choice provides the segment means for each covariate as well as segment distributions associated with 5 grouped levels (from low to high) of these variables. Notice that segment 1, the segment having the highest utility for 'no minimum balance' requirement, maintains a much lower balance than the other segments (a mean of $577.71 compared to $1,138.96 for segment 4 and higher for the other segments). Segment 4, the segment with the highest utility for free ATM use, has utilized ATMs more than the other segments. Similarly, segment 3, the segment most influenced by the 'no cost per check' option, writes more checks than the others.

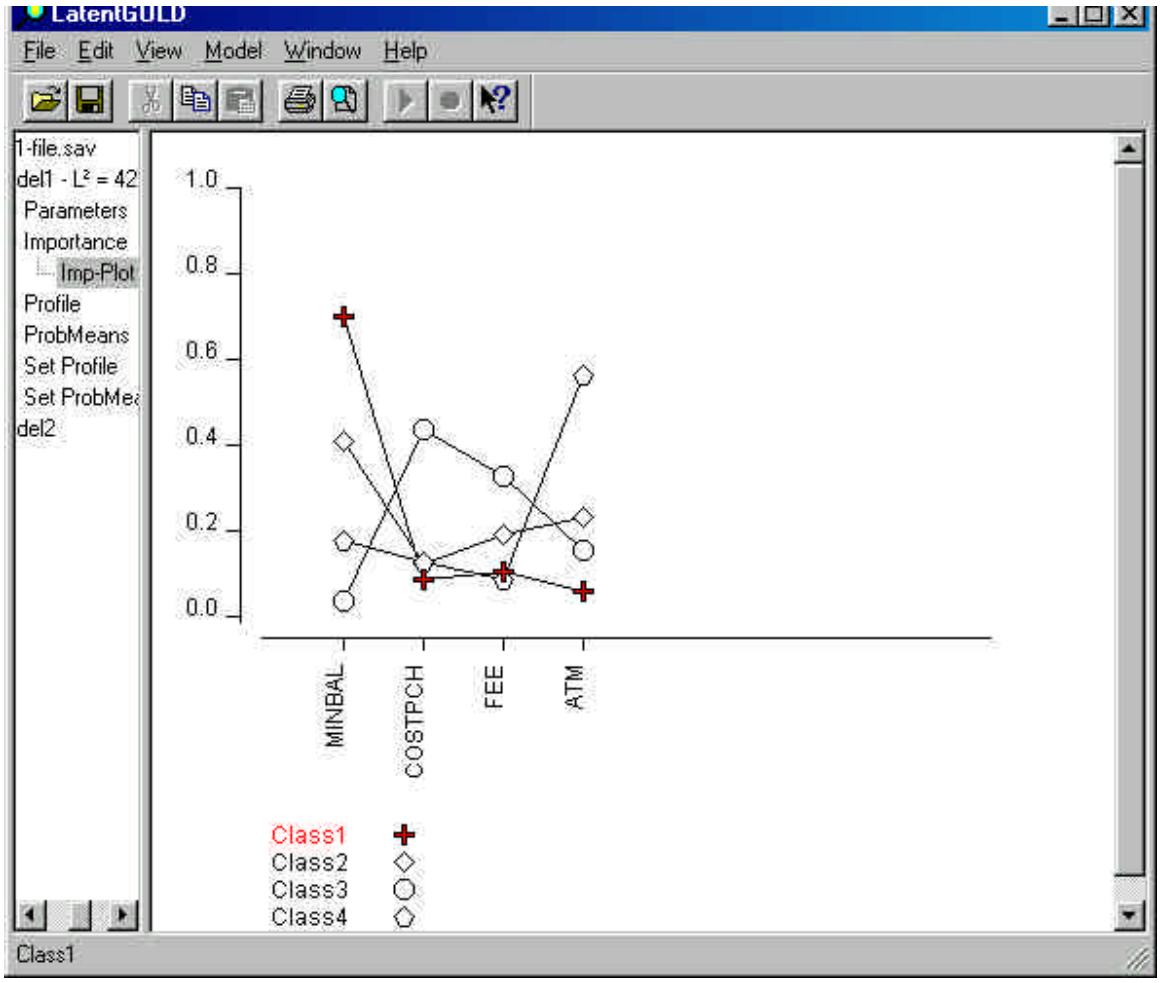
These differential preferences can be seen even more clearly by examining the Importance output

- Click on Importance to display the Importance Output file



	Class1	Class2	Class3	Class4
Maximum Effects				
MINBAL	9.9603	2.6239	0.3208	1.9366
COSTPCH	1.4138	0.8562	2.6798	1.4177
FEE	1.6922	1.2795	2.0356	1.0194
ATM	1.0729	1.5170	1.0151	5.8169
Relative Effects				
MINBAL	0.7044	0.4180	0.0530	0.1900
COSTPCH	0.1000	0.1364	0.4428	0.1391
FEE	0.1197	0.2039	0.3364	0.1000
ATM	0.0759	0.2417	0.1678	0.5708

- Click on expand icon (+) to the left of Importance
- Click on Imp-Plot to display the Importance Plot
- Click on 'Class1' below the plot to view graphically how segment 1 differs from the other segments in Importance



See Kamakura et. al. for similar interpretations of these segments. The numbering of these segments differs from the article because Latent Gold Choice assigns segment numbers by segment size. The correspondence between Latent Gold Choice segments and segments described in the Kamakura et. al. article is:

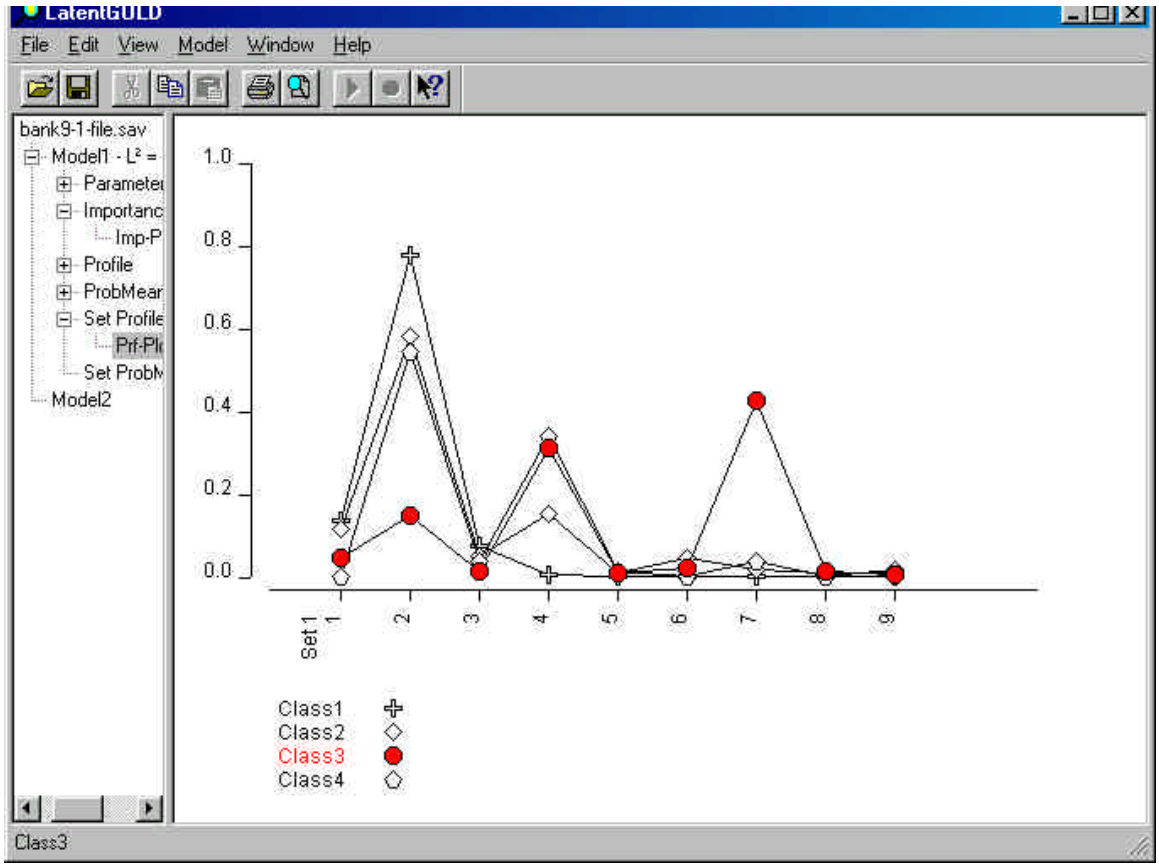
Latent GOLD Choice → Kamakura et. al.
1 → 4
2 → 3
3 → 1
4 → 2

We will now examine some further graphical output.

For each segment, the Set Profile Output presents the predicted probability of the highest ranked alternative. To display these probabilities graphically,

- Click on expand icon (+) to the left of Set Profile
- Click on Prf-Plot to display the Set Profile Plot
- Click on Class3 at the bottom of the plot

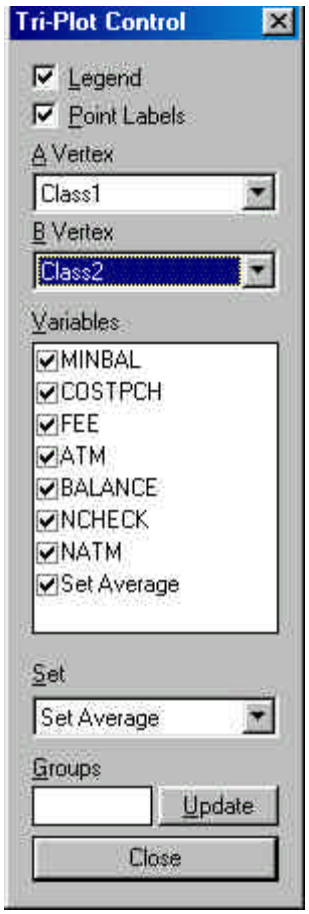
Notice that segment 3 is much more likely to rank alt#7 highest (the predicted probability for doing this is .43) than the other segments. The corresponding probability is less than .05 for each of the other segments.



- Click on expand icon (+) to the left of ProbMeans
- Click on Tri-Plot which contrasts 3 groups of segments

By default, the tri-plot contrasts segment 1, segment 2, and the combined segments 3 and 4. Since segments 3 and 4 are very different, it would be more informative to group segments 1 and 2. To customize the tri-plot so that the upper vertex corresponds to segments 1 and 2,

- Right-click on the tri-plot to retrieve the Plot control

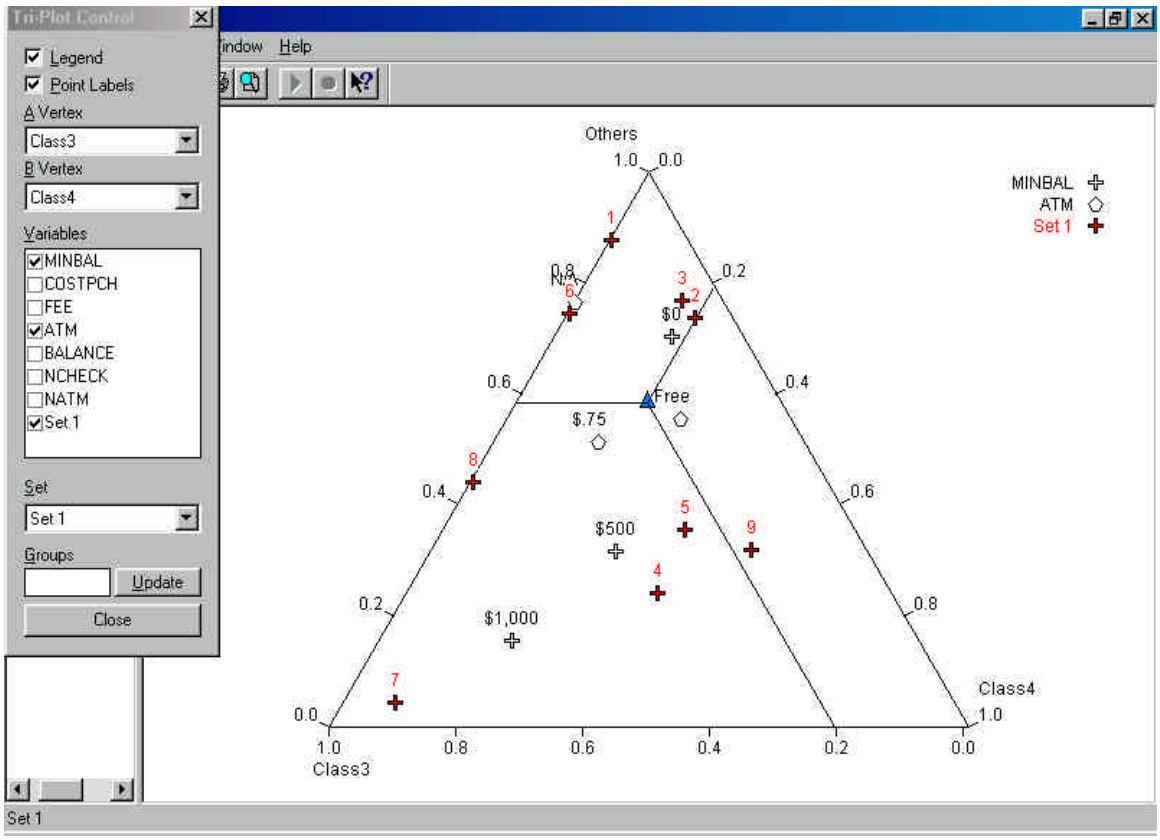


- In the A Vertex box, replace Class1 with Class3
- In the B Vertex box, replace Class2 with Class4

To make the plot more readable,

- click to left of each covariates (BALANCE, NCHECK & NATM) and to the left of the attributes COSTPCH and FEE to remove them from the display
- In the Set box, replace Average Set with Set 1 (This is not really necessary in this example because there is only a single set so Set 1 and Average Set are the same here)
- Click on the symbol for Set 1 in the upper right of the plot to highlight the first choice on the plot

Your screen should now look like this:

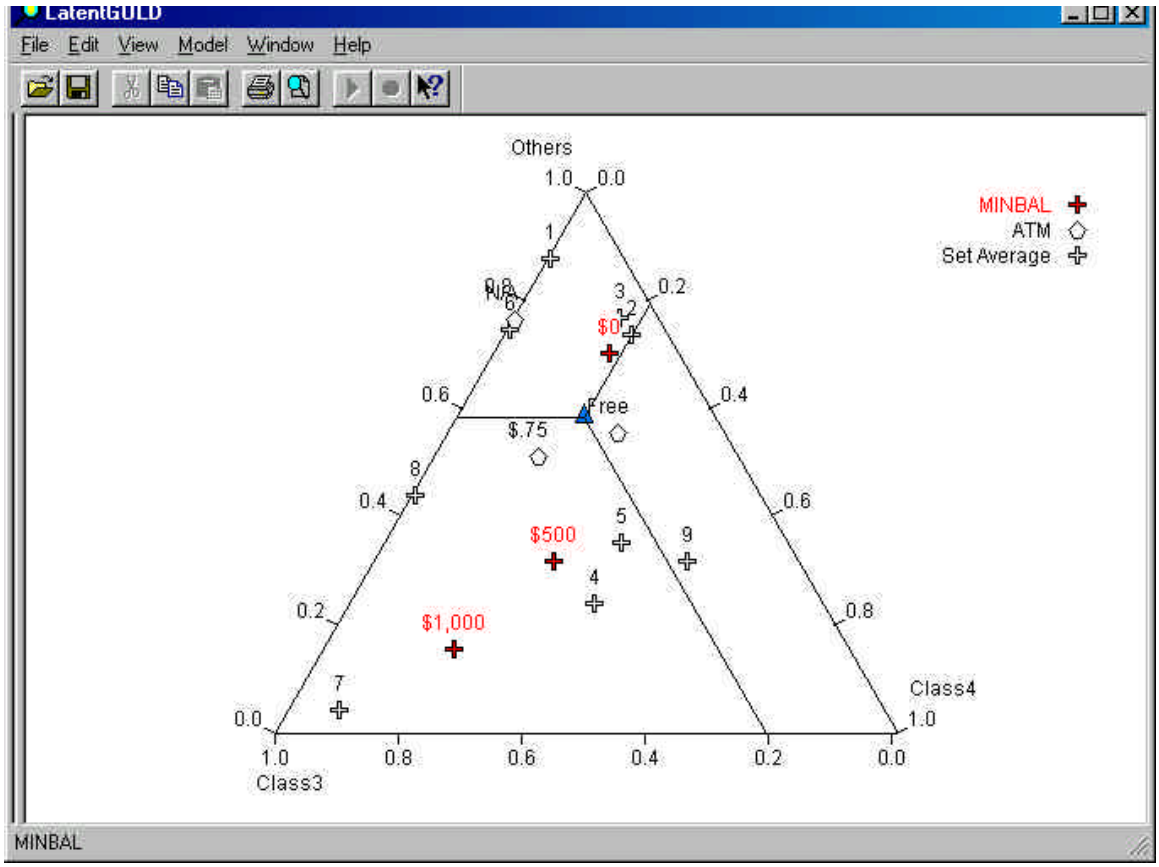


Notice that alt#7 is closest to the lower left vertex associated with segment 3. Alt#4 is near the bottom of the plot, associated more with segments 3 and 4 than 1 and 2. In contrast, segments 1 and 2 are more likely to select alts# 1 and 2 than the other segments.

To highlight the attribute levels in the plot

- Click on the label for MINBAL
- Click on the label for ATM

For example, when we highlight the levels of MINBAL, the proximity of the $1,000 minimum balance level to the lower left vertex of the following graph makes clear that segment 3 is more likely than the other segments to choose an account requiring a minimum balance of $1,000. Similarly, the proximity of the minimum balance level of $0 to upper vertex shows that segments 1 and 2 are least likely to choose accounts requiring a $1,000 minimum balance.

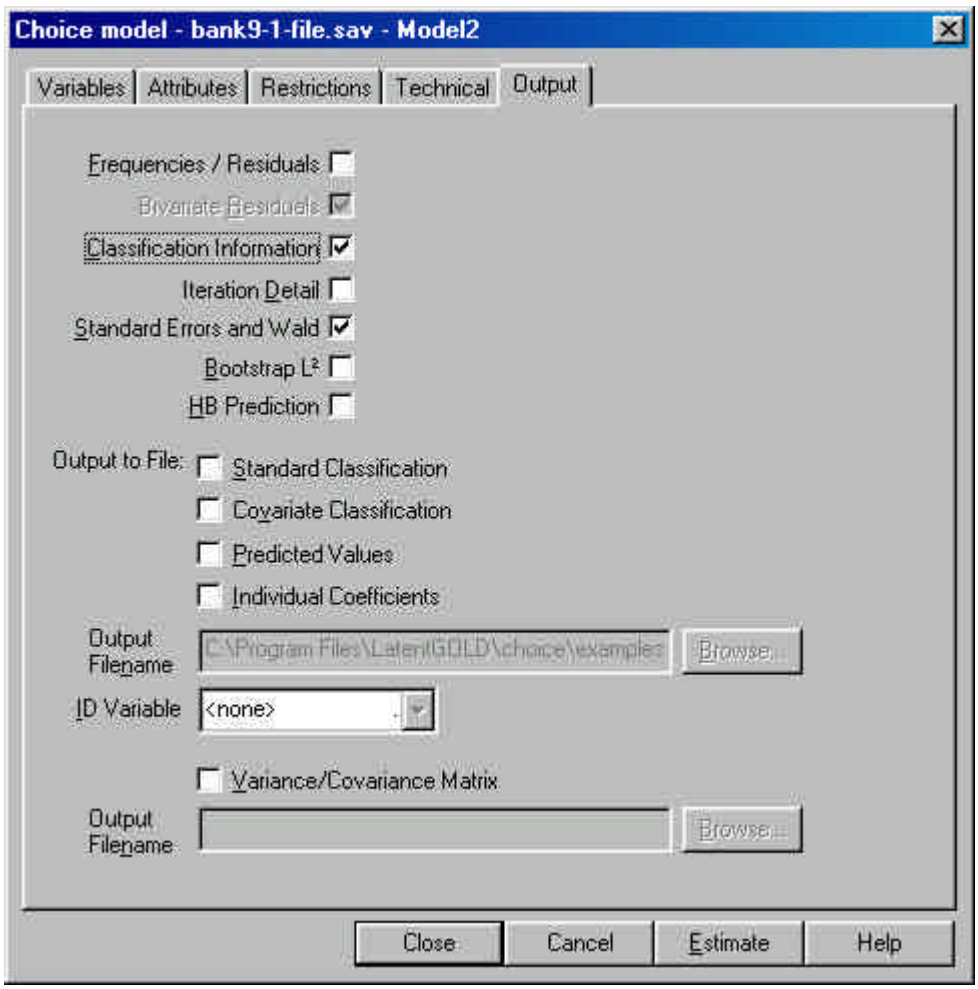


Classification output

Let's now turn to classifying respondents into the appropriate segments. For each respondent, the classification tables show the posterior membership probability associated with each class as well as the segment, along with the segment for which this probability is highest (displayed in the column marked ‘Modal’). The standard classification output takes into account responses as well as covariate information, while the covariate classification output uses only covariate information to compute these probabilities.

To view the standard classification output,

- On the left-hand screen, click on Model1
- Click on Output to open the output tab in the Model Analysis Dialog Box
- Click on Classification Information to request this output
- Click Estimate to re-estimate the model



After the estimation has completed,

- Click on Standard Classification in the Outline Pane

ID	BALANCE	CHECKS	ATM	Modal	Class1	Class2	Class3	Class4
1	80	3	5	1	0.9953	0.0047	0.0000	0.0000
2	482	35	8	2	0.2767	0.7232	0.0000	0.0001
3	395	11	4	2	0.0000	0.9976	0.0000	0.0024
4	198	16	8	1	0.9971	0.0029	0.0000	0.0000
5	350	19	0	1	0.9945	0.0055	0.0000	0.0000
6	218	12	2	2	0.0004	0.9994	0.0002	0.0000
7	44	2	0	2	0.0000	1.0000	0.0000	0.0000
8	3435	13	1	3	0.0000	0.0550	0.9450	0.0000
9	90	12	0	1	0.9596	0.0404	0.0000	0.0000
10	1687	7	0	3	0.0000	0.0334	0.9666	0.0000
11	329	2	1	1	0.9970	0.0030	0.0000	0.0000
12	756	20	1	1	0.6441	0.3558	0.0001	0.0001
13	1929	5	0	1	0.9828	0.0172	0.0000	0.0000
14	4823	10	8	4	0.0000	0.0011	0.0242	0.9747

- Click on Covariate classification

BALANCE	CHECKS	ATM	Modal	Class1	Class2	Class3	Class4
80	3	5	1	0.5239	0.1857	0.0785	0.2119
482	35	8	1	0.4192	0.2396	0.2190	0.1221
395	11	4	1	0.4593	0.2286	0.1308	0.1813
198	16	8	1	0.4883	0.2060	0.1100	0.1957
350	19	0	1	0.5046	0.2179	0.1782	0.0992
218	12	2	1	0.5296	0.2056	0.1266	0.1381
44	2	0	1	0.5840	0.1808	0.0905	0.1446
3435	13	1	3	0.0223	0.3764	0.4461	0.1552
90	12	0	1	0.5818	0.1871	0.1220	0.1091
1687	7	0	2	0.1675	0.3568	0.2743	0.2014
329	2	1	1	0.4915	0.2170	0.1100	0.1815
756	20	1	1	0.3777	0.2679	0.2325	0.1219
1929	5	0	2	0.1303	0.3705	0.2829	0.2163
4823	10	8	3	0.0037	0.3647	0.3895	0.2421

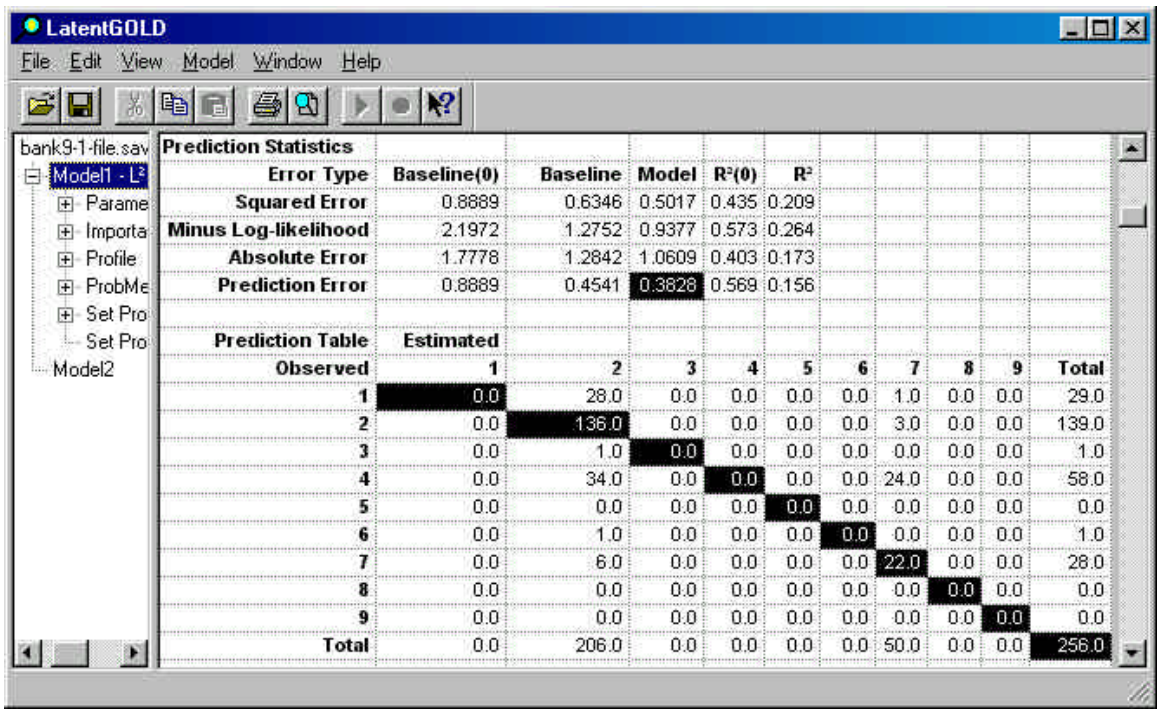
Now we will examine how well the model predicts the alternative that is ranked highest

- Click on Model 1
- Scroll down to display the prediction table

Prediction Table	Estimated									
Observed	1	2	3	4	5	6	7	8	9	Total
1	0.0	28.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	29.0
2	0.0	136.0	0.0	1.0	0.0	0.0	2.0	0.0	0.0	139.0
3	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
4	0.0	34.0	0.0	4.0	0.0	0.0	20.0	0.0	0.0	58.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
7	0.0	5.0	0.0	3.0	0.0	0.0	20.0	0.0	0.0	28.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	0.0	205.0	0.0	8.0	0.0	0.0	43.0	0.0	0.0	256.0

Notice that except for the 2 respondents who made other choices, the numbers in the right-most ‘Total’ column show that only 4 of the 9 alternatives were ranked highest. Specifically, we see that alt#1 was ranked highest by n=9 respondents, alt#2 by n=139), alt#4 by n=58 and alt#7 by the remaining 28. In our next tutorial, we will show how we can exploit this fact in model estimation .

- Now, scroll up to the Prediction Statistics



Prediction Statistics

Error Type	Baseline(0)	Baseline	Model	$R^2(0)$	R^2
Squared Error	0.8889	0.6346	0.5017	0.435	0.209
Minus Log-likelihood	2.1972	1.2752	0.9377	0.573	0.264
Absolute Error	1.7778	1.2842	1.0609	0.403	0.173
Prediction Error	0.8889	0.4541	0.3828	0.569	0.156

Prediction Table

Observed \ Estimated	1	2	3	4	5	6	7	8	9	Total
1	0.0	28.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	29.0
2	0.0	136.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	139.0
3	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
4	0.0	34.0	0.0	0.0	0.0	0.0	24.0	0.0	0.0	58.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
7	0.0	6.0	0.0	0.0	0.0	0.0	22.0	0.0	0.0	28.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	0.0	206.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	256.0

We see that overall, the model provides a hit rate of only 1-.3828 = 62% for these first choices.

In our next tutorial, we will also show how to estimate some simpler *partial* rank models, and we will see that they predict better than this full rank model. Even without the use of covariates, will obtain hit rates of over 80%.

References

Kamakura, W., Wedel, M., Agrawal, J. 1994. *Concomitant* variable latent class models for conjoint analysis. *International Journal of Research in Marketing 11, 451-464.*