Tutorial #S1: Getting Started with LG-Syntax™

DemoData = 'conjoint.sav'

This tutorial introduces the use of the LG-Syntax module, an add-on to the Advanced version of Latent GOLD. In this tutorial we utilize the data which was also used in 'Tutorial #3: LC Regression with Repeated Measures.'

The Goal

Since it is quite easy to setup a GUI model in Latent GOLD using the LG4.5 Windows Menu system, it is often useful to begin with a GUI model containing the basic elements of the desired syntax model. This GUI model can then be converted to an initial syntax model automatically using the 'Generate Syntax' option from the 'Model' menu. The goal of this tutorial is to illustrate this process, as well as show how to modify LG-Equations to obtain additional models.

In this tutorial we will:

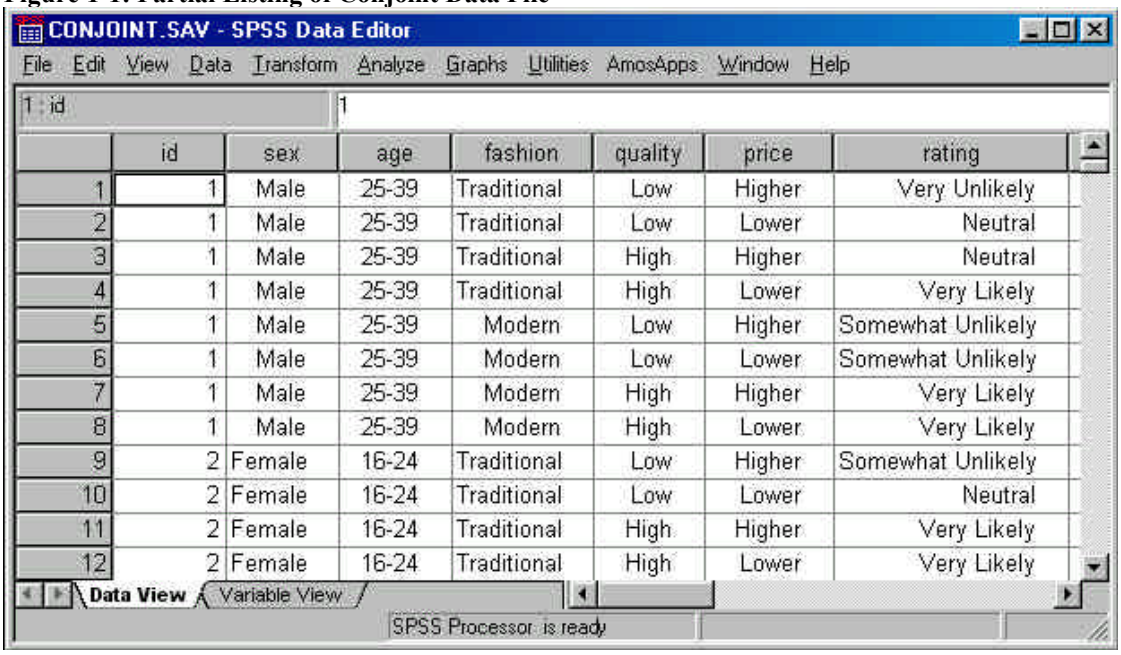
- Introduce the use of LG-Syntax
- Show how the LG-Syntax can be generated from a GUI model
- Examine the Equations section of the LG-Syntax
- Modify the LG-Equations to specify a different LC regression model
- See how parameter restrictions may be specified in different ways using the syntax

We will reuse the data from 'Tutorial 3: LC regression with Repeated Measures' in this tutorial. While the data were generated under the assumption of the ordinal logit model, for simplicity in introducing the equation section of the LG-Syntax we will treat the dependent variable as continuous rather than ordinal, so that the models obtained are LC (linear) regression models.

The Data

The data for this example are obtained from a hypothetical conjoint marketing study involving repeated measures where respondents were asked to provide likelihood of purchase ratings under each of several different scenarios. A partial listing of the data is shown in Figure 1-1.

Figure 1-1: Partial Listing of Conjoint Data File



	id	sex	age	fashion	quality	price	rating
1	1	Male	25-39	Traditional	Low	Higher	Very Unlikely
2	1	Male	25-39	Traditional	Low	Lower	Neutral
3	1	Male	25-39	Traditional	High	Higher	Neutral
4	1	Male	25-39	Traditional	High	Lower	Very Likely
5	1	Male	25-39	Modern	Low	Higher	Somewhat Unlikely
6	1	Male	25-39	Modern	Low	Lower	Somewhat Unlikely
7	1	Male	25-39	Modern	High	Higher	Very Likely
8	1	Male	25-39	Modern	High	Lower	Very Likely
9	2	Female	16-24	Traditional	Low	Higher	Somewhat Unlikely
10	2	Female	16-24	Traditional	Low	Lower	Neutral
11	2	Female	16-24	Traditional	High	Higher	Very Likely
12	2	Female	16-24	Traditional	High	Lower	Very Likely

As suggested in Figure 1-1, there are 8 records for each case (there are 400 cases in total); one record for each cell in this 2x2x2 complete factorial design of different scenarios for the purchase of a product:

- *FASHION (1 = Traditional; 2 = Modern)*
- *QUALITY (1 = Low; 2 = High)*
- *PRICE (1 = Lower; 2 = Higher)*

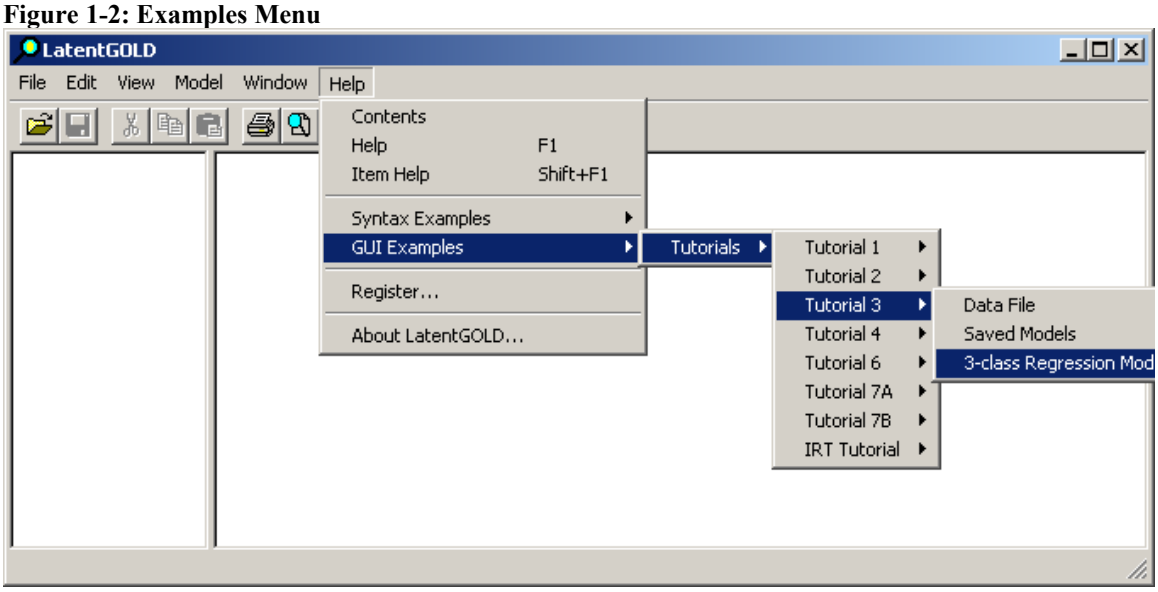
The dependent variable (*RATING*) is a rating of purchase intent on a five-point scale. The three attributes listed above will be used as predictor variables in the model. We will also include the two demographic variables as covariates, in a second model.

- *SEX (1 = Male; 2 = Female)*
- *AGE (1 = 16-24; 2 = 25-39; 3 = 40+).*

Using a GUI Example to Setup the Syntax Model

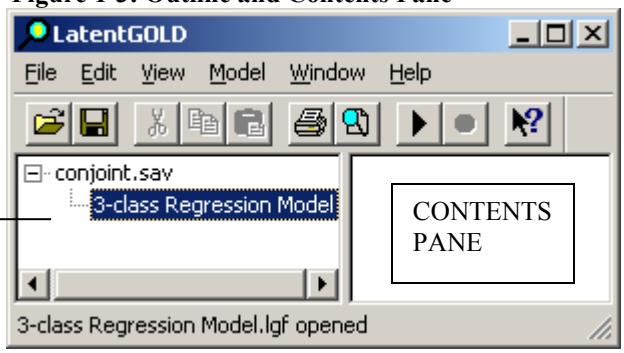
- Navigate as in Figure 1-2 to open the Example tutorial #3 GUI model named '3-class Regression Model'.

Figure 1-2: Examples Menu



As shown in Figure 1-3, the Outline pane contains the name of the data file along with the previously saved model(s). The Contents Pane (currently empty) shows the contents of selected model output.

Figure 1-3: Outline and Contents Pane



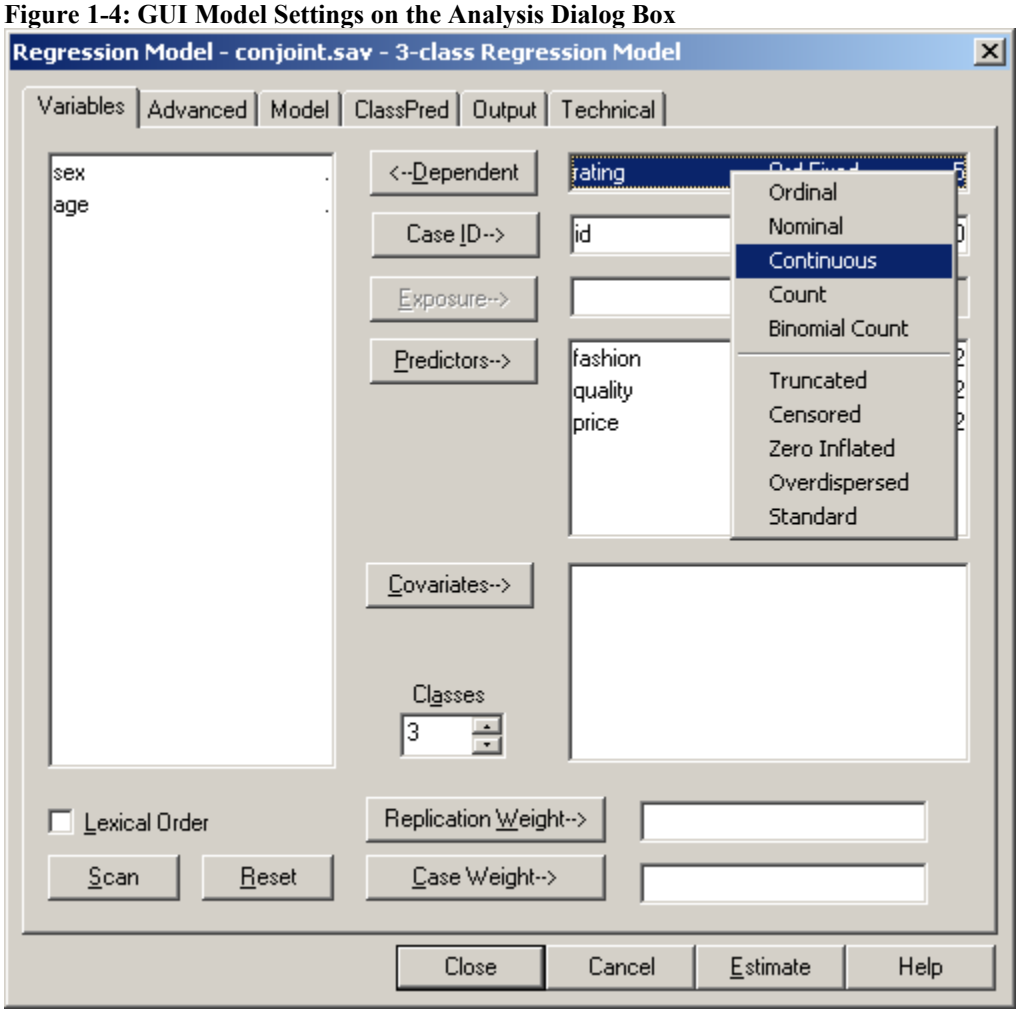
OUTLINE PANE

- Double click to open the Analysis Dialog Box for this model.

For reasons explained earlier in "The Goal" section, we will now change the scale type from Ordinal ("Ord-Fixed") to 'Continuous' to specify a (3-class) *linear* regression model.

- Right click on the dependent variable 'rating' to view the available scale types
- Select 'Continuous'(see Figure 1-4)

Figure 1-4: GUI Model Settings on the Analysis Dialog Box



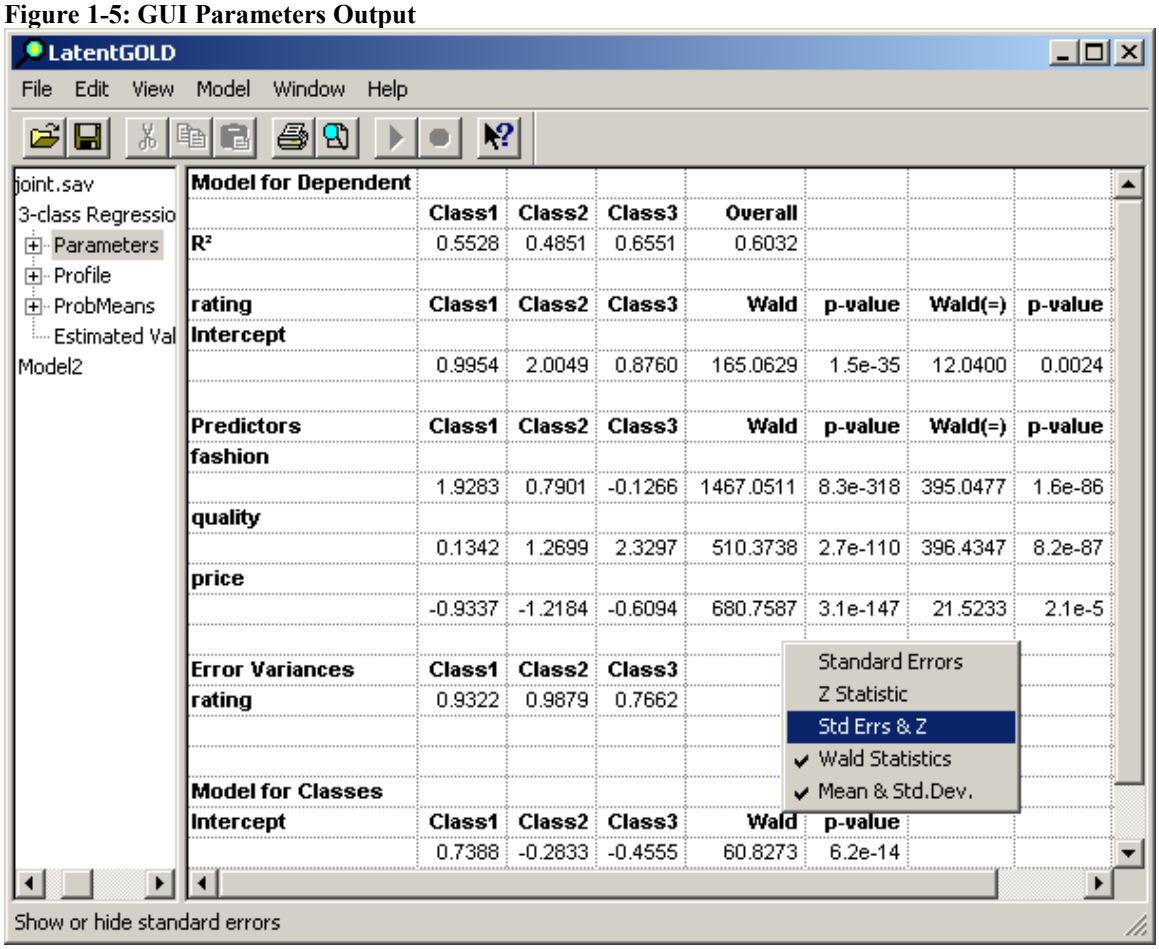
- Select 'Estimate' to estimate this gui model

  A Warning Message appears alerting you to the fact that the dependent variable contains fewer than 20 values.

- Select 'OK' to estimate the linear regression model anyway, despite the small number of values for the dependent variable.

- Click on 'Parameters' to view the Parameters output in the Contents Pane (see Figure 1-5)

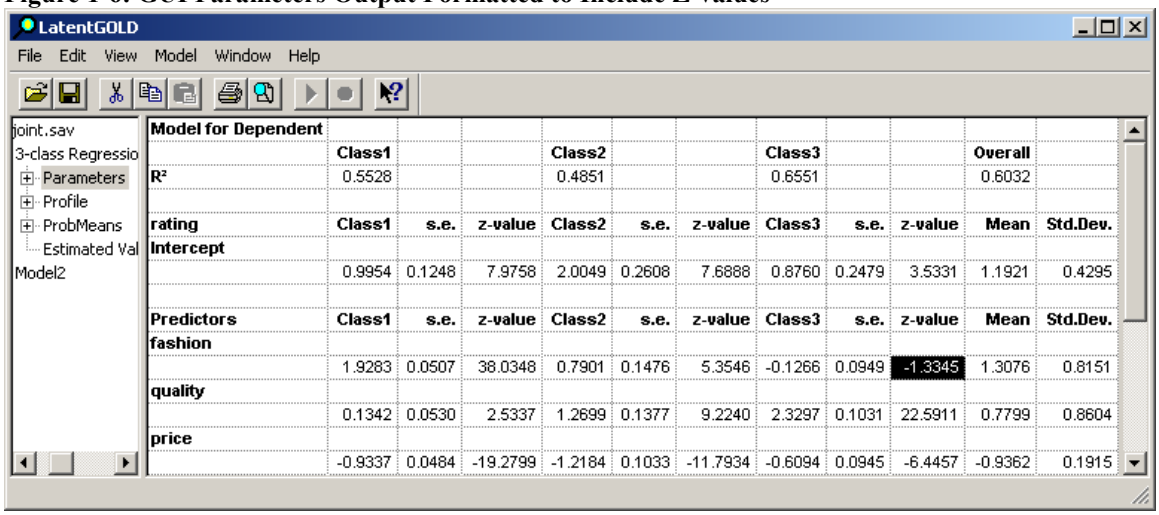
Figure 1-5: GUI Parameters Output



- Right click to retrieve the popup menu (shown above) and select 'Std Errs & Z' to display these statistics

Notice that the regression coefficient for FASHION is not significant for class 3 (as highlighted in Figure 1-6, $|Z| < 2$).

Figure 1-6: GUI Parameters Output Formatted to Include Z-values



- Double click on the model name '3-class Regression Model' or the new model named 'Model2' to re-open the Analysis Dialog Box (the new model is a copy of the last estimated model)
- Click on 'Model' to open the Model Tab
- Right click on the '3' associated with the FASHION coefficient for latent class 3, and select 'No effect' from the Pop-up menu to restrict this effect to 0 (see Figure 1-7).
- Click Estimate to estimate this restricted model

Figure 1-7: GUI Model Tab

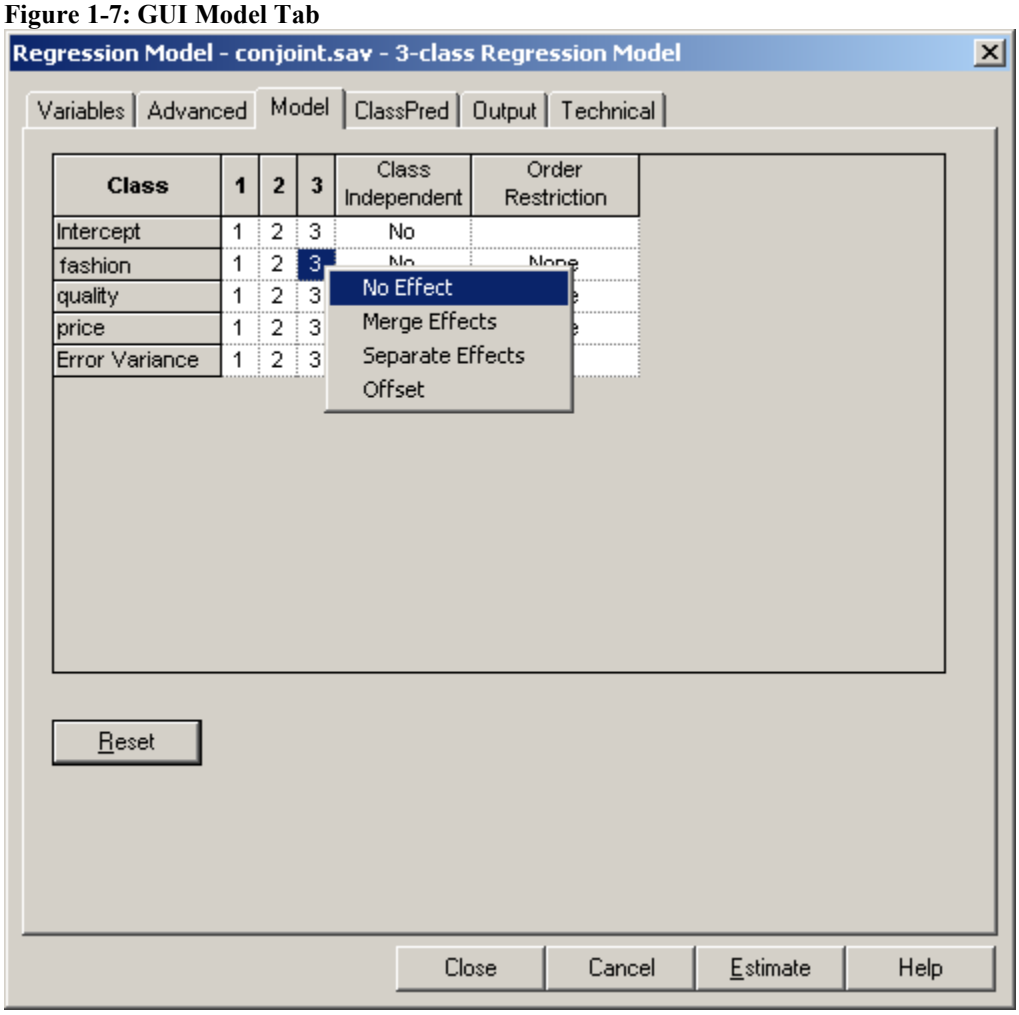
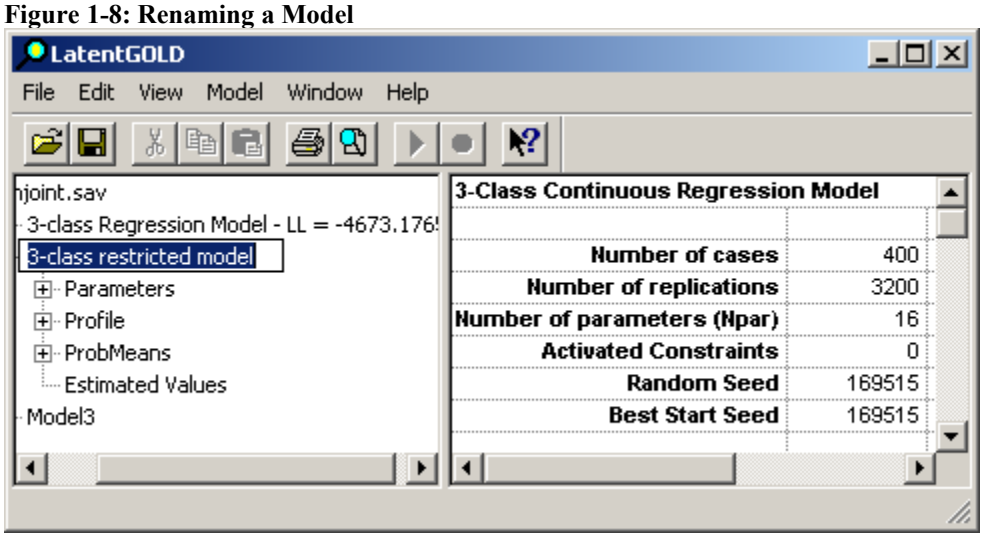


Figure 1-8: Renaming a Model



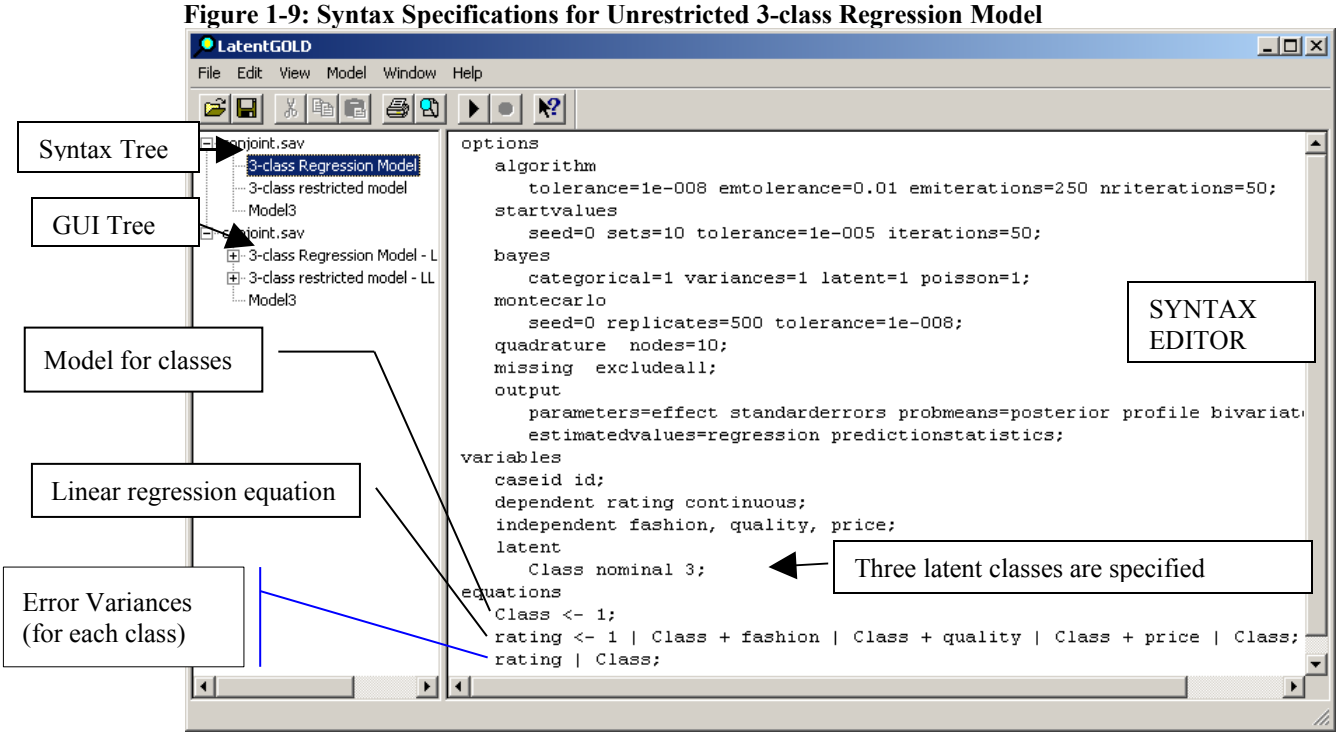
- Click on the model name ('Model2') to select it, click again to enter the edit model, and type '3-class restricted model' to rename it (see Figure 1-8).

To generate the syntax specifications for all models listed in the Outline Pane for the Conjoint.sav' file:

- Select the file name 'Conjoint.sav'
- From the Model Menu, select 'Generate syntax'

A separate (new) syntax tree appears above the GUI tree as shown in Figure 1-9, similar to the appearance when you open a different dataset in the GUI. Note that the GUI model names are preserved in the syntax. The first of these models, the unrestricted model, is highlighted and the syntax specifications for it appear in Contents Pane. It can be edited.

Figure 1-9: Syntax Specifications for Unrestricted 3-class Regression Model



- At the bottom of the syntax three equations appear, each ending with a ';'. The special keyword '1' denotes the intercepts.

  - The first equation corresponds to the Model for the Classes, which contains only the intercepts. These intercepts are logit parameters, which yield the class size probabilities.

  - The second is the linear regression equation with terms separated by '+', coefficients to be estimated for each such term. The skeletal structure of this model is expressed by 'rating = 1 + fashion + quality + price', with conditional effects (' | Class') specified for each effect. 'Conditional effects' means that separate coefficients are estimated for each latent class.

  - The third equation specifies that separate error variances are estimated for each class.

Estimating a Syntax Model

To estimate this model:

- From the Model menu choose 'Estimate'

or, you may also select the  on the toolbar

Upon completion of the estimation, the log-likelihood (LL) appears in the Outline Pane to the right of the model name (see Figure 1-10), and the syntax and output listings appear as separate entries in the expanded syntax tree. The model output listings in this expanded syntax tree appear somewhat differently than in the corresponding version of the expanded GUI tree (recall Figure 1-5):

- The syntax statements appear as an additional entry named 'syntax'

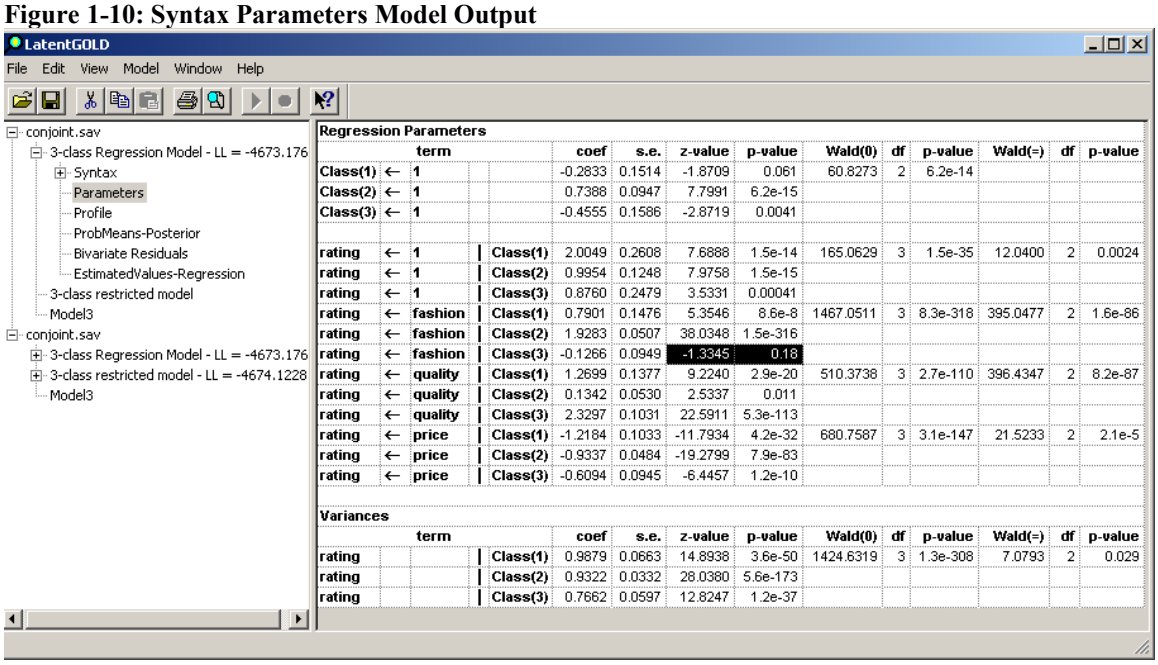
- The ProbMeans output is named 'ProbMeans-Posterior' to distinguish it from other versions of the ProbMeans that are available in the syntax.
- The Estimated Values output is named 'EstimatedValues-Regression' to distinguish it from other versions of the Estimated Values that are available in the syntax.
- The Bivariate Residuals output appears as an additional entry because it is available (by default) in the syntax but not available at all in the GUI for regression models.

The equivalence of the unrestricted models estimated from the syntax and GUI can be confirmed by comparing the LL values and verifying that they are equal (LL = -4673.173).

➢ Select 'Parameters' to view the Parameters output in the Contents Pane

Notice that the Parameters output is formatted differently than the GUI. Among the differences, a p-value column is present in the syntax output, showing that the Z-value of -1.3 (highlighted in Figure 1-10) is not significant at the .05 level (p=.18).

Figure 1-10: Syntax Parameters Model Output



Regression Parameters

		term		coef	s.e.	z-value	p-value	Wald(0)	df	p-value	Wald(=)	df	p-value
Class(1)	←	1		-0.2833	0.1514	-1.8709	0.061	60.8273	2	6.2e-14			
Class(2)	←	1		0.7388	0.0947	7.7991	6.2e-15						
Class(3)	←	1		-0.4555	0.1586	-2.8719	0.0041						
rating	←	1	Class(1)	2.0049	0.2608	7.6888	1.5e-14	165.0629	3	1.5e-35	12.0400	2	0.0024
rating	←	1	Class(2)	0.9954	0.1248	7.9758	1.5e-15						
rating	←	1	Class(3)	0.8760	0.2479	3.5331	0.00041						
rating	←	fashion	Class(1)	0.7901	0.1476	5.3546	8.6e-8	1467.0511	3	8.3e-318	395.0477	2	1.6e-86
rating	←	fashion	Class(2)	1.9283	0.0507	38.0348	1.5e-316						
rating	←	fashion	Class(3)	-0.1266	0.0949	-1.3345	0.18						
rating	←	quality	Class(1)	1.2699	0.1377	9.2240	2.9e-20	510.3738	3	2.7e-110	396.4347	2	8.2e-87
rating	←	quality	Class(2)	0.1342	0.0530	2.5337	0.011						
rating	←	quality	Class(3)	2.3297	0.1031	22.5911	5.3e-113						
rating	←	price	Class(1)	-1.2184	0.1033	-11.7934	4.2e-32	680.7587	3	3.1e-147	21.5233	2	2.1e-5
rating	←	price	Class(2)	-0.9337	0.0484	-19.2799	7.9e-83						
rating	←	price	Class(3)	-0.6094	0.0945	-6.4457	1.2e-10						

Variances

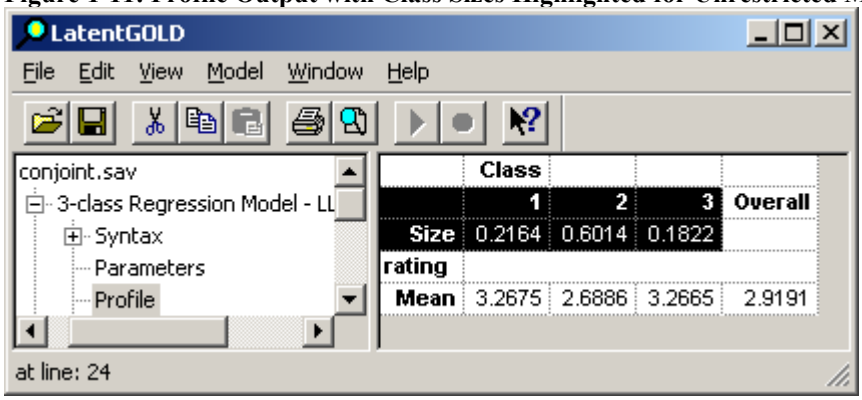
	term		coef	s.e.	z-value	p-value	Wald(0)	df	p-value	Wald(=)	df	p-value
rating		Class(1)	0.9879	0.0663	14.8938	3.6e-50	1424.6319	3	1.3e-308	7.0793	2	0.029
rating		Class(2)	0.9322	0.0332	28.0380	5.6e-173						
rating		Class(3)	0.7662	0.0597	12.8247	1.2e-37						

Unlike the results from a model estimated in the GUI, the latent classes in an unrestricted syntax model will not necessarily be ordered from high to low. Figure 1-11 (below) shows that in this particular run, latent class 2 is the largest class, while class 1 is the second largest. If we estimated the model again, the ordering may be different.

To view the classes sizes:

➢ Select 'Profile' in the syntax tree.

Figure 1-11: Profile Output with Class Sizes Highlighted for Unrestricted Model



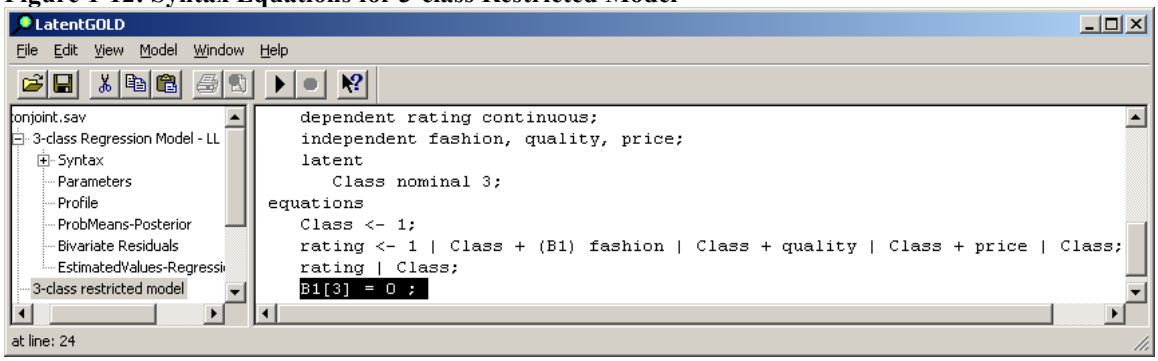
Restricting Certain Effects to be Zero or Class Independent

The syntax allows substantial flexibility in placing restrictions on the model parameters. To view the specification generated automatically from the restricted GUI model:

- Select the model name '3-class restricted model'

The syntax appears in the Contents Pane.

Figure 1-12: Syntax Equations for 3-class Restricted Model

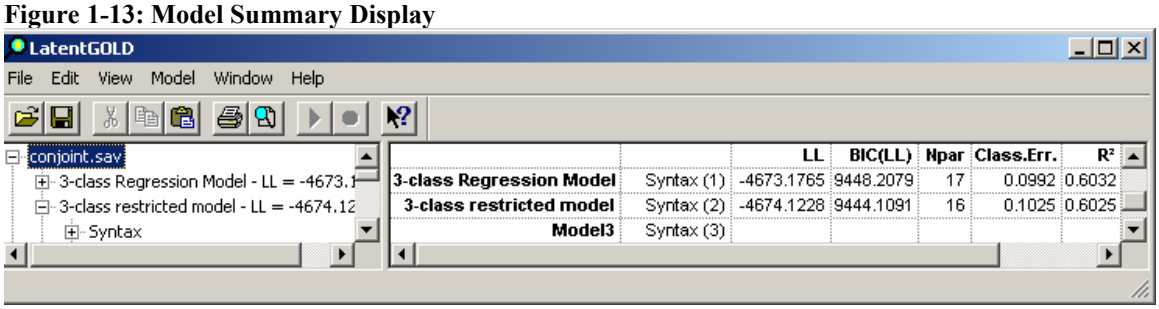


Notice that an additional equation appears that embodies the parameter restriction. B1(3) = 0 states that the parameter B1 for latent class 3 is set to 0. The B1 coefficients are defined in the regression equation by adding '(B1)' as part of the conditional effect term for the predictor FASHION. This defines 3 B1 parameters, one for each class. These parameters are referenced as B1[1], B1[2] and B1[3]. By restricting B1[3] = 0, this serves to define latent class 3 as the one for which the B1 effect is zero.

To estimate this model,

- select  on the toolbar
- Select the data file name to see how the models compare.

Figure 1-13: Model Summary Display

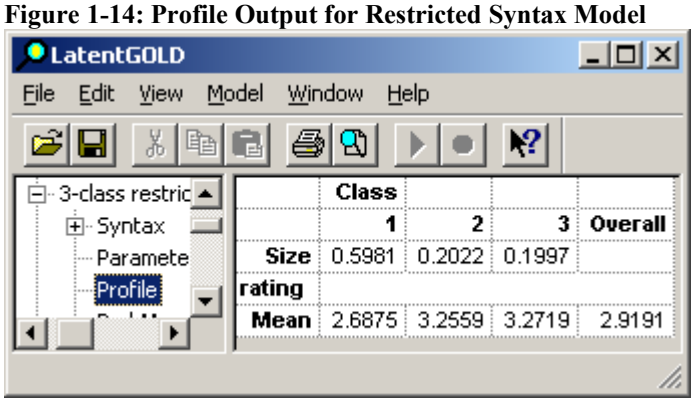


		LL	BIC(LL)	Npar	Class.Err.	R^2
3-class Regression Model	Syntax (1)	-4673.1765	9448.2079	17	0.0992	0.6032
3-class restricted model	Syntax (2)	-4674.1228	9444.1091	16	0.1025	0.6025
Model3	Syntax (3)					

The restricted model has one fewer parameter (Npar = 16 vs. 17 for the unrestricted model), is preferred according to the BIC (9444 vs. 9448) and both models have an $R^2 = .60$.

- Select 'Profile' to view the class sizes:

Figure 1-14: Profile Output for Restricted Syntax Model

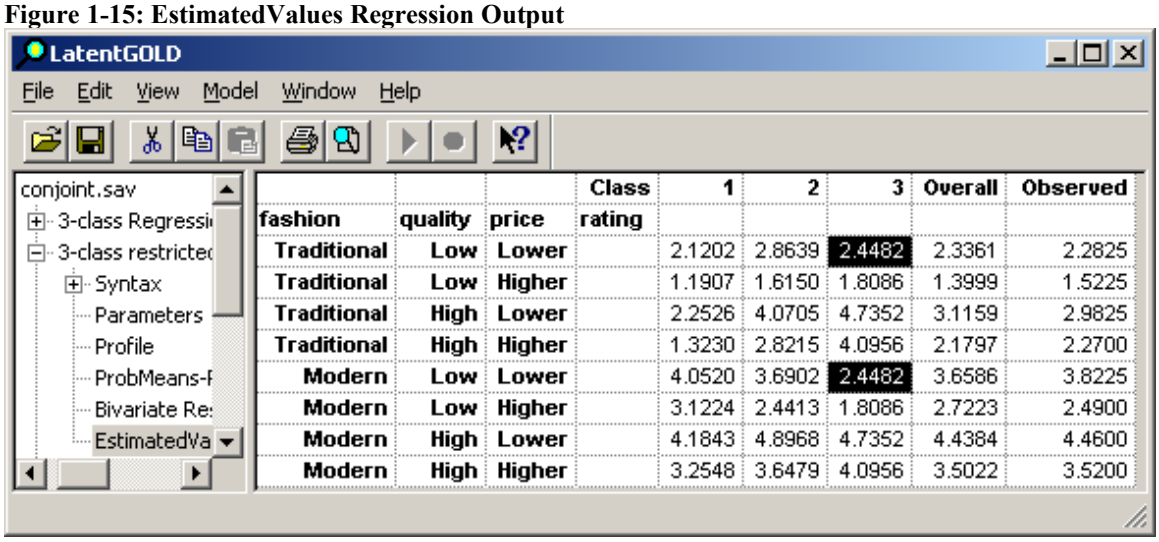


	Class			
	1	2	3	Overall
Size	0.5981	0.2022	0.1997	
rating				
Mean	2.6875	3.2559	3.2719	2.9191

The Figure above shows that the largest class is now the first class. If we re-estimated this restricted model again, classes 1 and 2 may be reversed, but class 3 will remain as class 3 because of the restriction. Comparing the Parameter output between the 2 models, you will find that the parameter estimates did not change much, if you keep in mind that the ordering of the classes may have changed.

- Select 'EstimatedValues-Regression' to view the expected ratings.

Figure 1-15: EstimatedValues Regression Output



			Class	1	2	3	Overall	Observed
fashion	quality	price	rating					
Traditional	Low	Lower		2.1202	2.8639	2.4482	2.3361	2.2825
Traditional	Low	Higher		1.1907	1.6150	1.8086	1.3999	1.5225
Traditional	High	Lower		2.2526	4.0705	4.7352	3.1159	2.9825
Traditional	High	Higher		1.3230	2.8215	4.0956	2.1797	2.2700
Modern	Low	Lower		4.0520	3.6902	2.4482	3.6586	3.8225
Modern	Low	Higher		3.1224	2.4413	1.8086	2.7223	2.4900
Modern	High	Lower		4.1843	4.8968	4.7352	4.4384	4.4600
Modern	High	Higher		3.2548	3.6479	4.0956	3.5022	3.5200

Since the ratings for class 3 are not affected by FASHION, we see that the expected rating for this class is the same when evaluating Traditional vs. Modern shoes of the same PRICE and QUALITY. This is not the case for classes 1 or 2.

Inclusion of Covariates in Model

Next, we will expand the model for the classes to include the covariates. Main effects for the covariates could be specified in the GUI, and a syntax model could have been generated as before. But here, we will use the syntax to specify both main effects and interaction effects for the covariates.

- Select 'Model 3' to view a copy of the syntax for the 3-class restricted model in the Syntax Editor.
- Scroll down to the Independent Variables Section.
- Type ", age, sex' to include the covariates AGE and SEX as additional independent variables (see Figure 1-16B). Alternatively, if you right click in the Contents Pane, you can retrieve a list of data file variables from which you can select the covariates and the variable names will be copied to the syntax (see Figure 1-16A)
- Type " + age + sex + age sex" to include the main effects and interaction effect of the covariates in the model for classes.
- Select  on the toolbar to estimate this model.

Figure 1-16A: List of Data File Variables

Figure 1-16B: Including Covariates in Syntax Editor

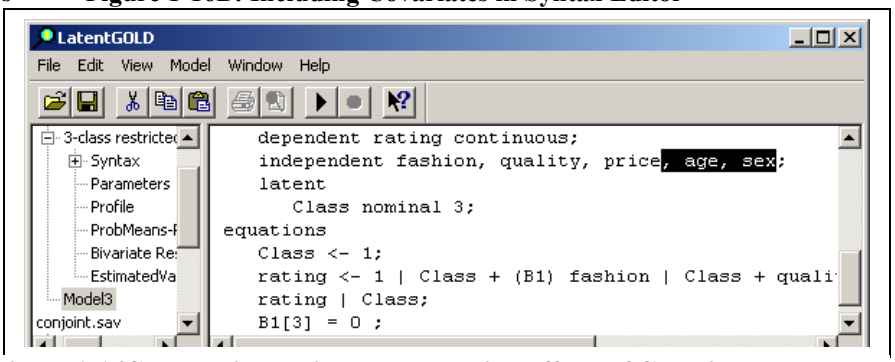
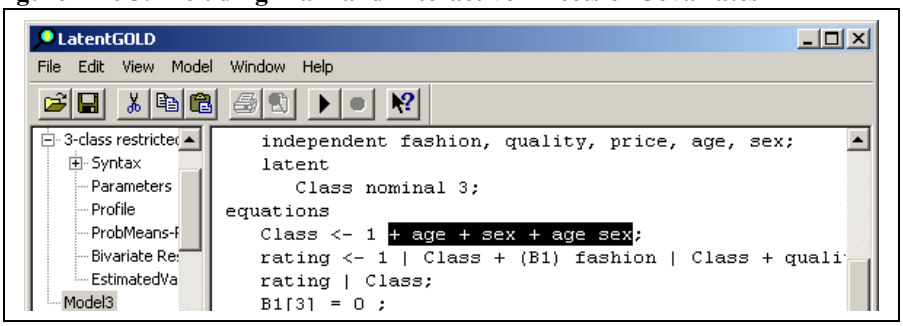
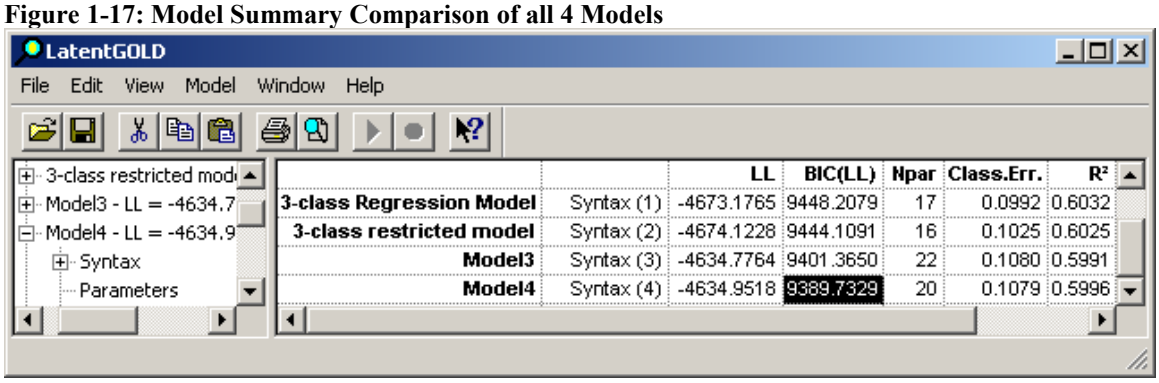


Figure 1-16C: Including Main and Interactive Effects of Covariates



- Select 'Model 4' to view a copy of the syntax for the last model estimated.
- Scroll down to the Model for classes.
- Remove the interaction effect.
- Select  on the toolbar to estimate this model.
- Select the data file to compare all models.

Figure 1-17: Model Summary Comparison of all 4 Models



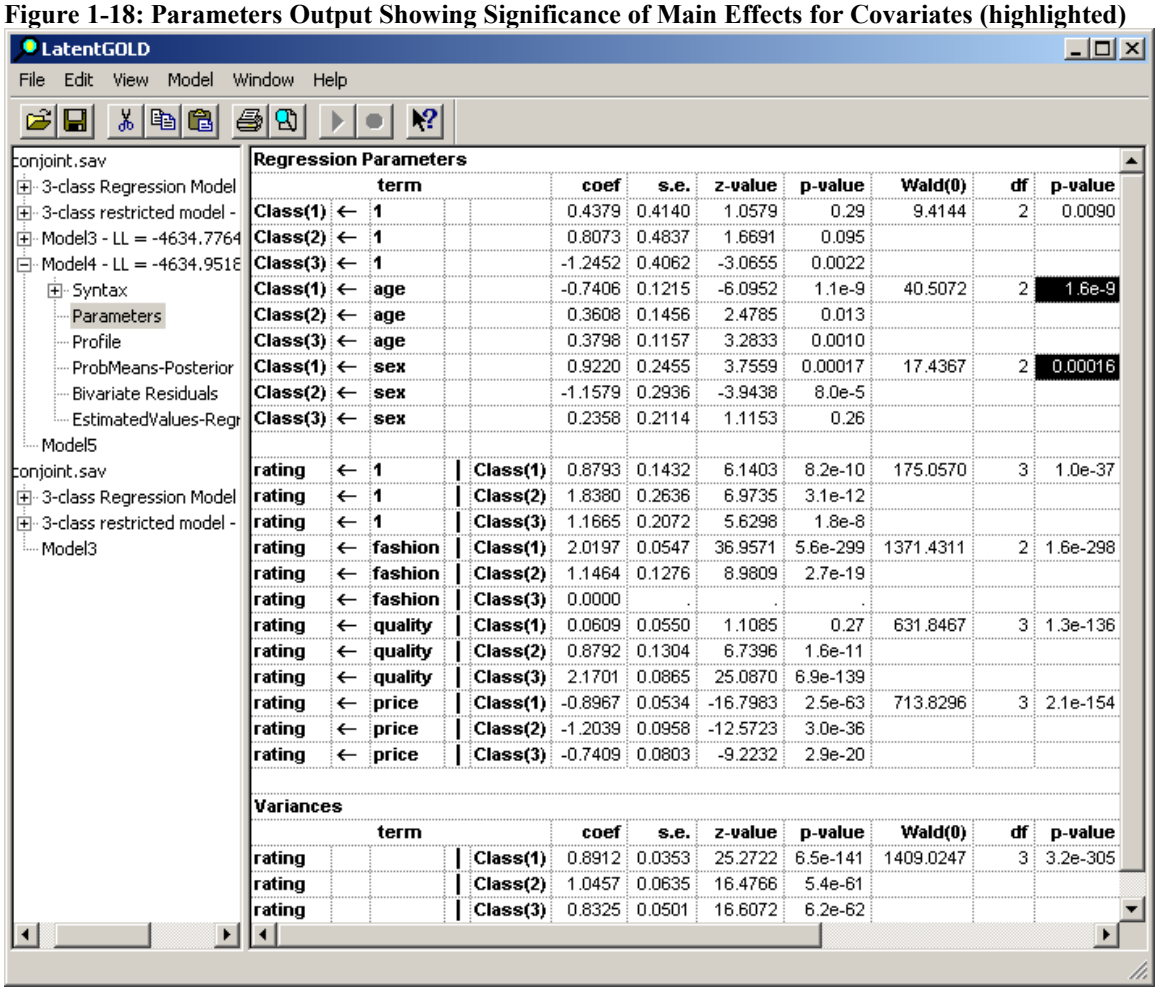
		LL	BIC(LL)	Npar	Class.Err.	R^2
3-class Regression Model	Syntax (1)	-4673.1765	9448.2079	17	0.0992	0.6032
3-class restricted model	Syntax (2)	-4674.1228	9444.1091	16	0.1025	0.6025
Model3	Syntax (3)	-4634.7764	9401.3650	22	0.1080	0.5991
Model4	Syntax (4)	-4634.9518	9389.7329	20	0.1079	0.5996

According to the BIC, both Models 3 and 4 are preferred over the models without these covariates. Model 4 provides the best fit suggesting that the interaction effect is not significant.

- Select 'Parameters' for Model 4 to display the Parameters output for this model.

Note that the covariates AGE and SEX are both significant in the Model4.

Figure 1-18: Parameters Output Showing Significance of Main Effects for Covariates (highlighted)



Regression Parameters

term				coef	s.e.	z-value	p-value	Wald(0)	df	p-value
Class(1) ←	1			0.4379	0.4140	1.0579	0.29	9.4144	2	0.0090
Class(2) ←	1			0.8073	0.4837	1.6691	0.095			
Class(3) ←	1			-1.2452	0.4062	-3.0655	0.0022			
Class(1) ←	age			-0.7406	0.1215	-6.0952	1.1e-9	40.5072	2	1.6e-9
Class(2) ←	age			0.3608	0.1456	2.4785	0.013			
Class(3) ←	age			0.3798	0.1157	3.2833	0.0010			
Class(1) ←	sex			0.9220	0.2455	3.7559	0.00017	17.4367	2	0.00016
Class(2) ←	sex			-1.1579	0.2936	-3.9438	8.0e-5			
Class(3) ←	sex			0.2358	0.2114	1.1153	0.26			
rating ←	1		Class(1)	0.8793	0.1432	6.1403	8.2e-10	175.0570	3	1.0e-37
rating ←	1		Class(2)	1.8380	0.2636	6.9735	3.1e-12			
rating ←	1		Class(3)	1.1665	0.2072	5.6298	1.8e-8			
rating ←	fashion		Class(1)	2.0197	0.0547	36.9571	5.6e-299	1371.4311	2	1.6e-298
rating ←	fashion		Class(2)	1.1464	0.1276	8.9809	2.7e-19			
rating ←	fashion		Class(3)	0.0000	.	.	.			
rating ←	quality		Class(1)	0.0609	0.0550	1.1085	0.27	631.8467	3	1.3e-136
rating ←	quality		Class(2)	0.8792	0.1304	6.7396	1.6e-11			
rating ←	quality		Class(3)	2.1701	0.0865	25.0870	6.9e-139			
rating ←	price		Class(1)	-0.8967	0.0534	-16.7983	2.5e-63	713.8296	3	2.1e-154
rating ←	price		Class(2)	-1.2039	0.0958	-12.5723	3.0e-36			
rating ←	price		Class(3)	-0.7409	0.0803	-9.2232	2.9e-20			

Variances

term		coef	s.e.	z-value	p-value	Wald(0)	df	p-value
rating	Class(1)	0.8912	0.0353	25.2722	6.5e-141	1409.0247	3	3.2e-305
rating	Class(2)	1.0457	0.0635	16.4766	5.4e-61			
rating	Class(3)	0.8325	0.0501	16.6072	6.2e-62			

- Select 'Bivariate Residuals' for Model4 to display these statistics

Note that the BVRs for both covariates are small, suggesting that the assumption that the relationship between the covariates and dependent variable is explained by the latent classes. The BVRs are not available in the GUI for regression models. If one or more of these BVRs were large, the covariate(s) with the large BVRs could be included also as a predictor in the regression model.

Figure 1-19: Bivariate Residuals – Regression Display for Model with Covariates