Tutorial #3A: Using Latent GOLD 5.0 with Holdout Records

Overview

In this tutorial, we revisit the conjoint data example from Tutorial 3. Recall that these data were simulated from a 3-class ordinal regression model. In Tutorial 3 we used the variable ID to identify 400 cases, each containing 8 records (repeated measures) which provide their ratings for 8 different products that differed with respect to 3 dichotomous attributes (FASHION, QUALITY and PRICE). In the current tutorial, we will use only 4 of these 8 replications to estimate the parameters. The other 4 replications will be treated as holdout records.

In the Appendix of this tutorial we show how to use saved parameters to score new cases.

The Data

Data file: 'CONJOINT2.sav'
Saved model: 'CONJOINT2.lgf'

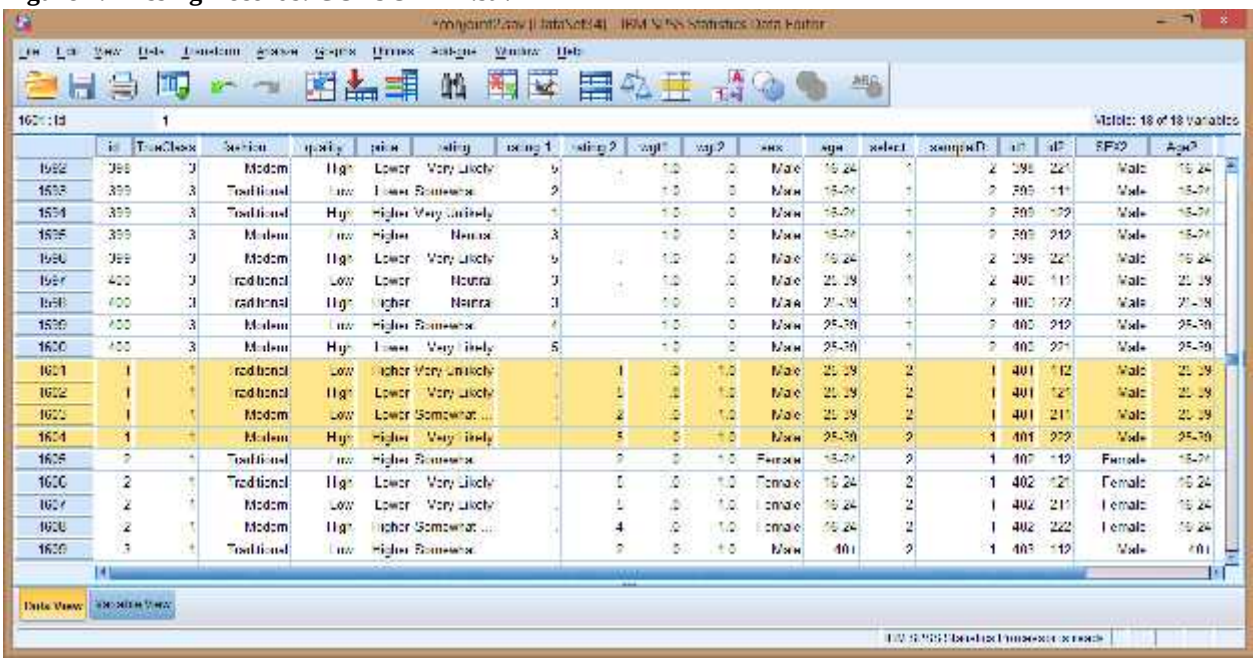
The file CONJOINT2.sav contains the same data as in the original file CONJOINT.sav as well as some additional variables such as TrueClass, which provides the true class membership for each case. As shown in Figure 1, a new ID variable named 'ID1' maintains the original ID#s 1-400 for the replication records that will be used for estimation, and 401-800 for the holdout records. The original dependent variable 'RATING' contains responses for all cases, while the new response variable called 'RATING.1' contains responses for cases ID1 = 1-400 only, responses for cases ID1 = 401-800 being replaced by missing values.

In this tutorial, use of RATING.1 as the dependent variable together with the missing values option 'Exclude Cases' causes the model parameters to be estimated using only half the replications. The resulting model will be used to score the replications associated with the remaining half of the replications (*holdout* records).

In Part 1 of this tutorial, we will utilize ID1 as the case ID variable, so that these holdout records, which are not used in the estimation, will be treated as responses from new cases (*holdout cases*).

In Part 2 we repeat our analysis using 'ID' as the case ID variable, thus treating the holdouts as holdout responses from the same cases. As we will see, in this latter situation, class membership for the 400 cases will be based on the 4 responses used to estimate the model, while in the former situation class membership will be based on the available covariates AGE and SEX.

Figure 1: Missing Records: CONJOINT2.sav



Part 1: Holdout Cases

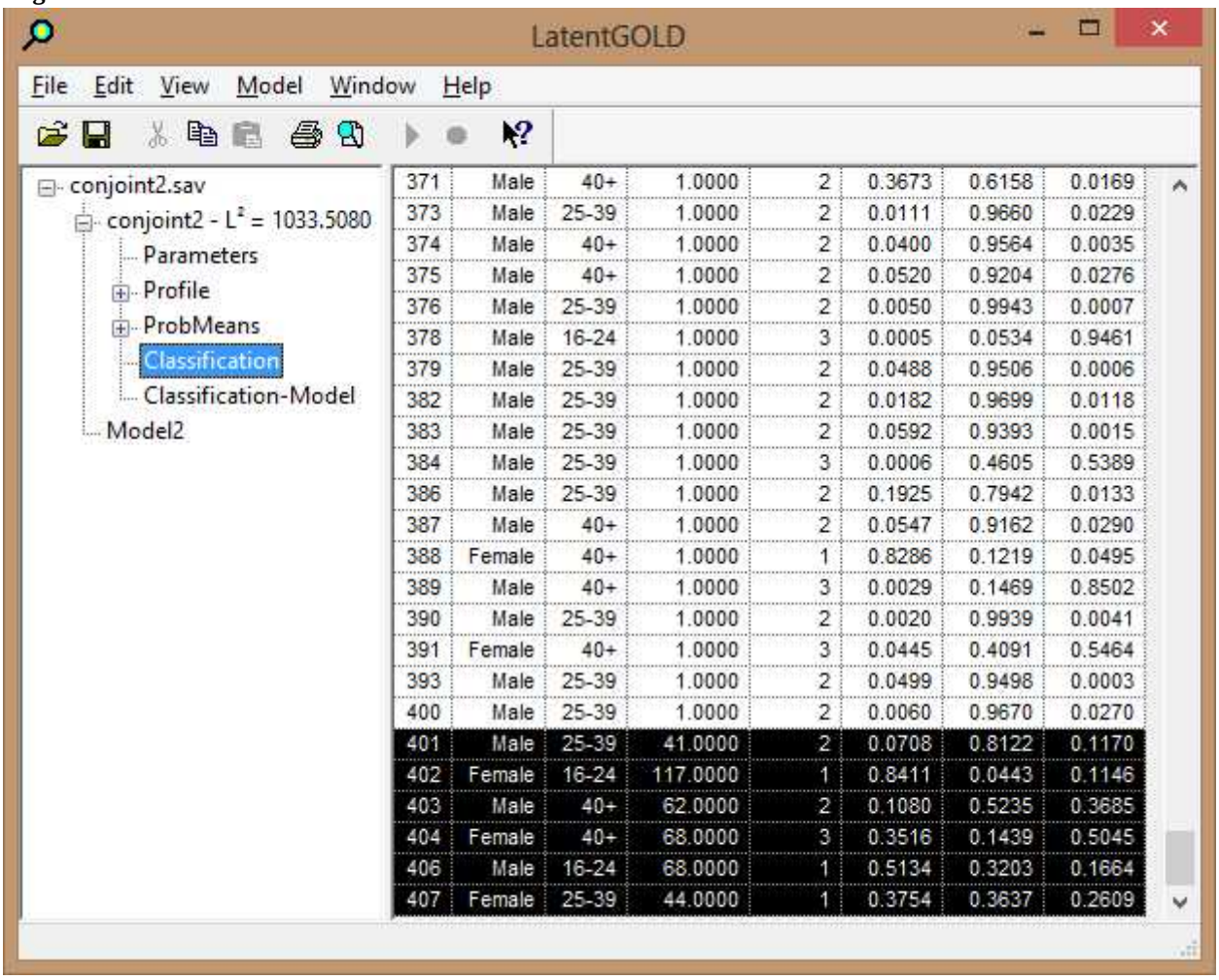
- Use the File menu in Latent GOLD to open CONJOINT2.lgf.
- Double click on CONJOINT2 in the Outline Pane to open the Analysis Dialog window.
- Click SCAN in the lower-left hand corner of the window.
- Click ESTIMATE to estimate a 3-class model.

Viewing Output and Interpreting Results

- Click to view the Classification output (Figure 2).
- Scroll down to view the new cases (ID1 values of 401 and greater).

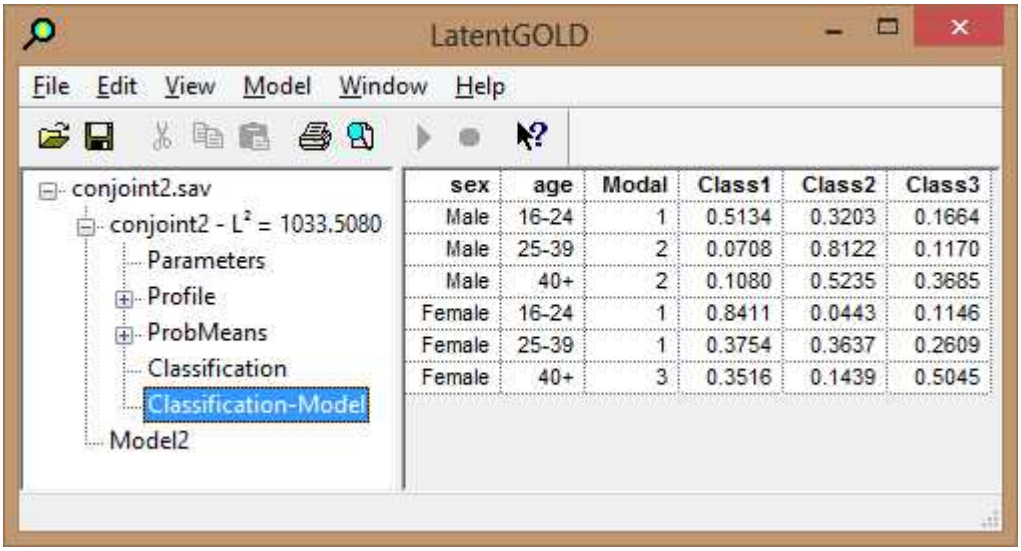
Since no response was used in the model estimation for the cases, they are assigned to the three latent classes solely based on their values on the covariates AGE and GENDER. Notice that the posterior membership probabilities are the same as given in the Covariate Classification output (Figure 3).

Figure 2: 'Classification – Posterior'



371	Male	40+	1.0000	2	0.3673	0.6158	0.0169
373	Male	25-39	1.0000	2	0.0111	0.9660	0.0229
374	Male	40+	1.0000	2	0.0400	0.9564	0.0035
375	Male	40+	1.0000	2	0.0520	0.9204	0.0276
376	Male	25-39	1.0000	2	0.0050	0.9943	0.0007
378	Male	16-24	1.0000	3	0.0005	0.0534	0.9461
379	Male	25-39	1.0000	2	0.0488	0.9506	0.0006
382	Male	25-39	1.0000	2	0.0182	0.9699	0.0118
383	Male	25-39	1.0000	2	0.0592	0.9393	0.0015
384	Male	25-39	1.0000	3	0.0006	0.4605	0.5389
386	Male	25-39	1.0000	2	0.1925	0.7942	0.0133
387	Male	40+	1.0000	2	0.0547	0.9162	0.0290
388	Female	40+	1.0000	1	0.8286	0.1219	0.0495
389	Male	40+	1.0000	3	0.0029	0.1469	0.8502
390	Male	25-39	1.0000	2	0.0020	0.9939	0.0041
391	Female	40+	1.0000	3	0.0445	0.4091	0.5464
393	Male	25-39	1.0000	2	0.0499	0.9498	0.0003
400	Male	25-39	1.0000	2	0.0060	0.9670	0.0270
401	Male	25-39	41.0000	2	0.0708	0.8122	0.1170
402	Female	16-24	117.0000	1	0.8411	0.0443	0.1146
403	Male	40+	62.0000	2	0.1080	0.5235	0.3685
404	Female	40+	68.0000	3	0.3516	0.1439	0.5045
406	Male	16-24	68.0000	1	0.5134	0.3203	0.1664
407	Female	25-39	44.0000	1	0.3754	0.3637	0.2609

Figure 3: Covariate Classification

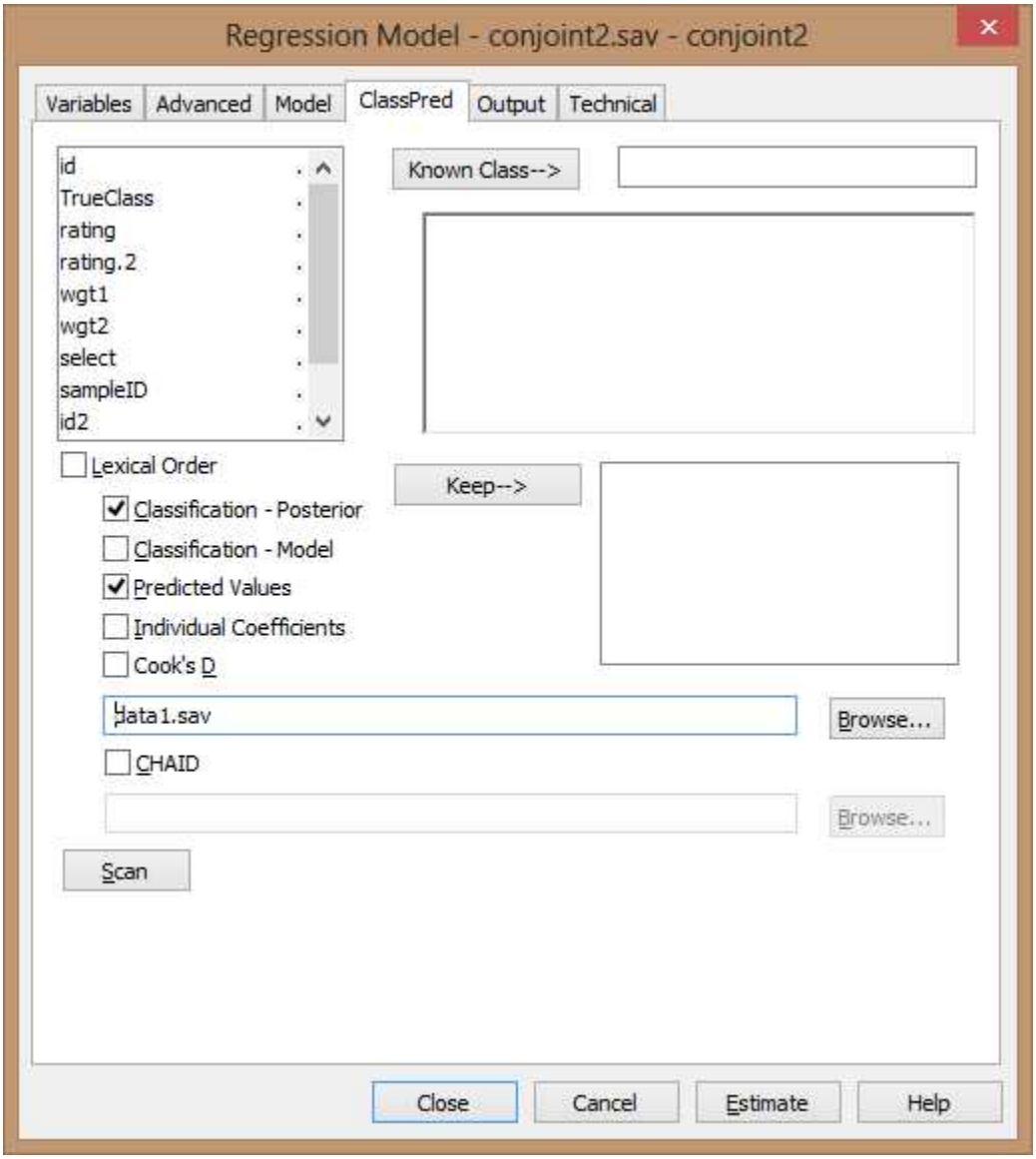


sex	age	Modal	Class1	Class2	Class3
Male	16-24	1	0.5134	0.3203	0.1664
Male	25-39	2	0.0708	0.8122	0.1170
Male	40+	2	0.1080	0.5235	0.3685
Female	16-24	1	0.8411	0.0443	0.1146
Female	25-39	1	0.3754	0.3637	0.2609
Female	40+	3	0.3516	0.1439	0.5045

To see the classification information and predictions for each case, we will now generate such output to a file.

- Double click on the model 'CONJOINT2' in the Outline Pane to open the Analysis Dialog window.
- In the ClassPred tab, check the boxes for 'Classification – Posterior' and Predicted Values. Note the directory where the output SPSS file is saved in the Output Filename box (Figure 4).
- Click ESTIMATE.

Figure 4: ClassPred

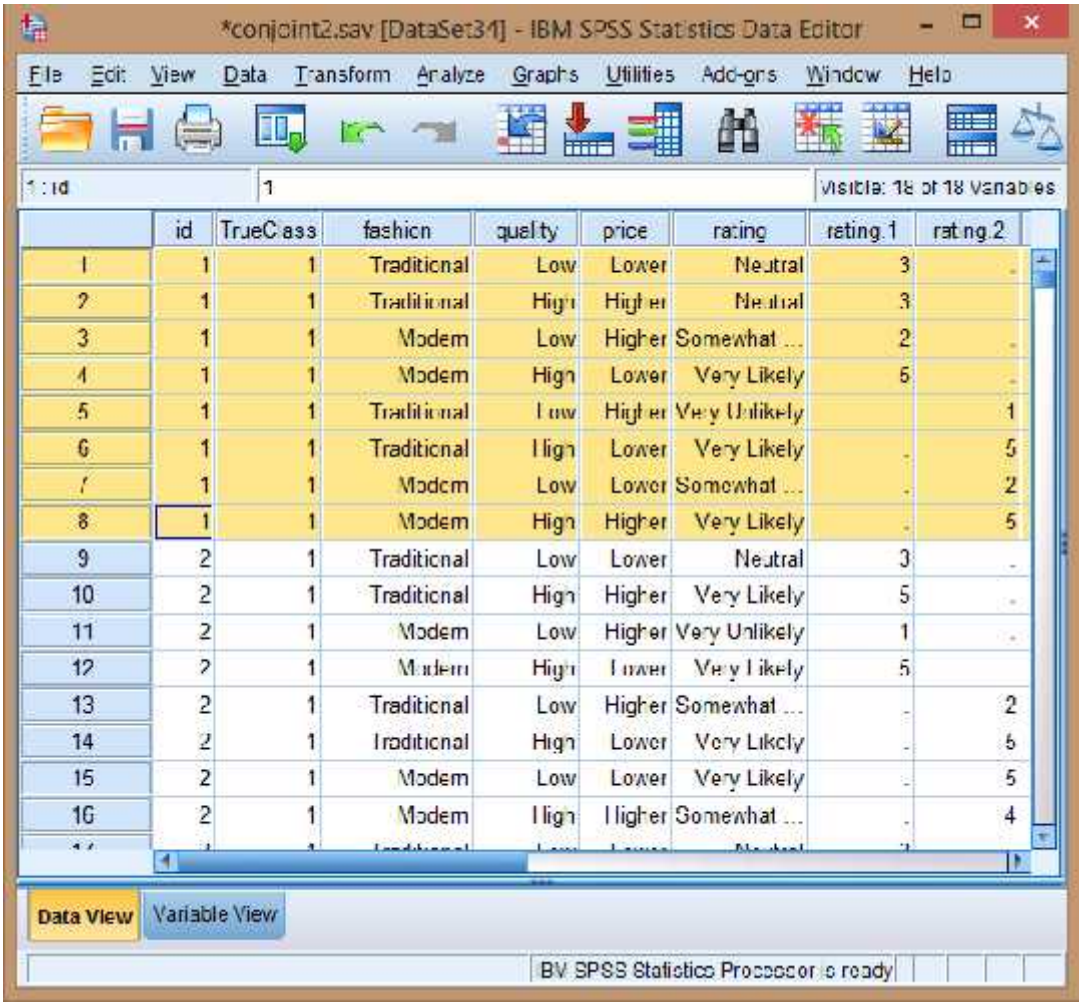


Note that Latent GOLD will output ‘Classification – Posterior’ values and Predicted Values for the missing cases as well.

Part 2: Holdout Replications for each Case

Now we will repeat the analysis, but this time, we will assume that the records containing responses are additional replications from the same 400 cases (Figure 5).

Figure 5: Missing Records from Same Case: CONJOINT2.sav



- Double click on Model3 to open the Analysis Dialog window.
- Click on 'ID1' in the Case ID box, and click on the CASE ID button to move 'ID1' to the Variables box.
- Click on 'ID' in the Variables box and click on the CASE ID button to move 'ID' to the CASE ID box.
- Click Estimate.
- Click to view the Parameters output.

Comparing these with the parameters estimated in our earlier model we see that the model is exactly the same as before because we're using the same information as before to estimate the model – namely, only the records containing the non-missing responses.

- Click to view Classification output.

We see that classifications are now different because each case is assigned to a class based on the 4 non-missing responses for that case.

- Double click on Model3 in the Outline Pane to open the Analysis Dialog window.
- In the ClassPred tab, check the boxes for 'Classification – Posterior' and Predicted Values. Note the directory where the output SPSS file is saved.
- Click ESTIMATE.

The predictions associated with the missing response records are now different than before because the posterior membership probabilities for these cases are different than before. (Recall from section 7.1.5 of the Technical Guide that predictions are generated for each case by taking a weighted average of the class-level predictions, using the posterior membership probabilities for that case as the weights).

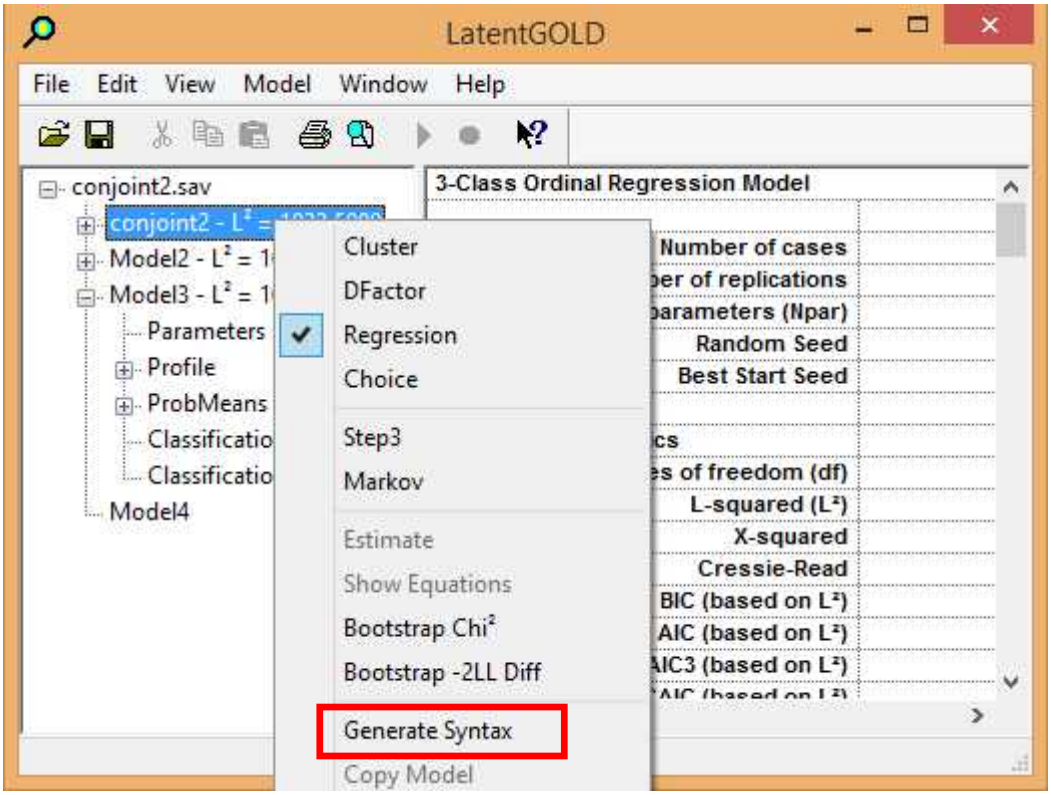
Appendix: Using the LG-Syntax module to Score New Cases

Alternatively, the parameter estimates can be saved and used to the score new cases.

We will begin by generating the syntax for the model 'conjoint2':

- Right click on the model 'conjoint2' in the Outline (left hand) pane.
- Select 'Generate syntax' (see Figure 6).

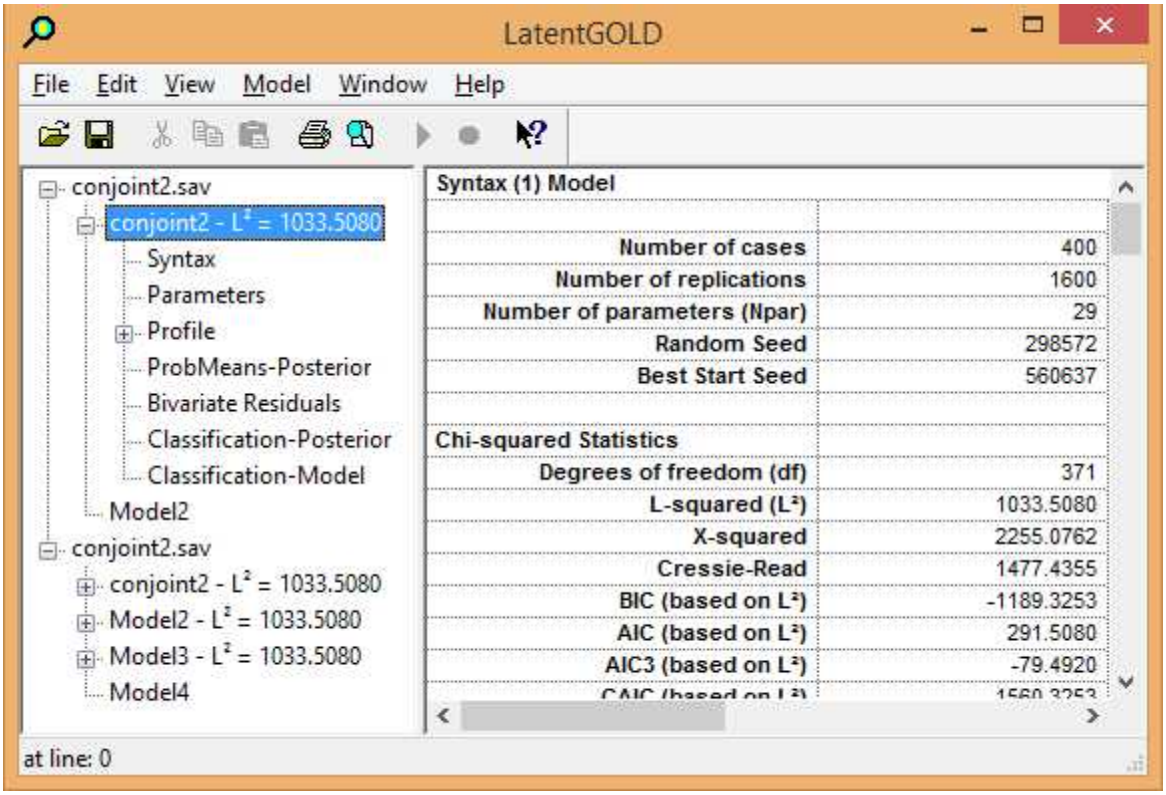
Figure 6. Generating Syntax.



A separate (new) syntax tree appears above the other tree.

- From the Syntax tree, right click on 'conjoint2' and click 'Estimate'.
- Verify that the $L^2 = 1033.5080$ (the same as we obtained in the GUI as shown in Figure 2).

Figure 7. Syntax model.



Next, we will save the syntax model with the saved model parameters as a .lgs file:

- From the Syntax tree, click on ‘conjoint2’.
- Click File→Save Syntax.
- From the ‘Save Contents’ drop down menu, select ‘with Parameters’ to save the model parameters (see Figure 8).
- Click ‘Save’

Figure 8. Saving a syntax model with parameters.

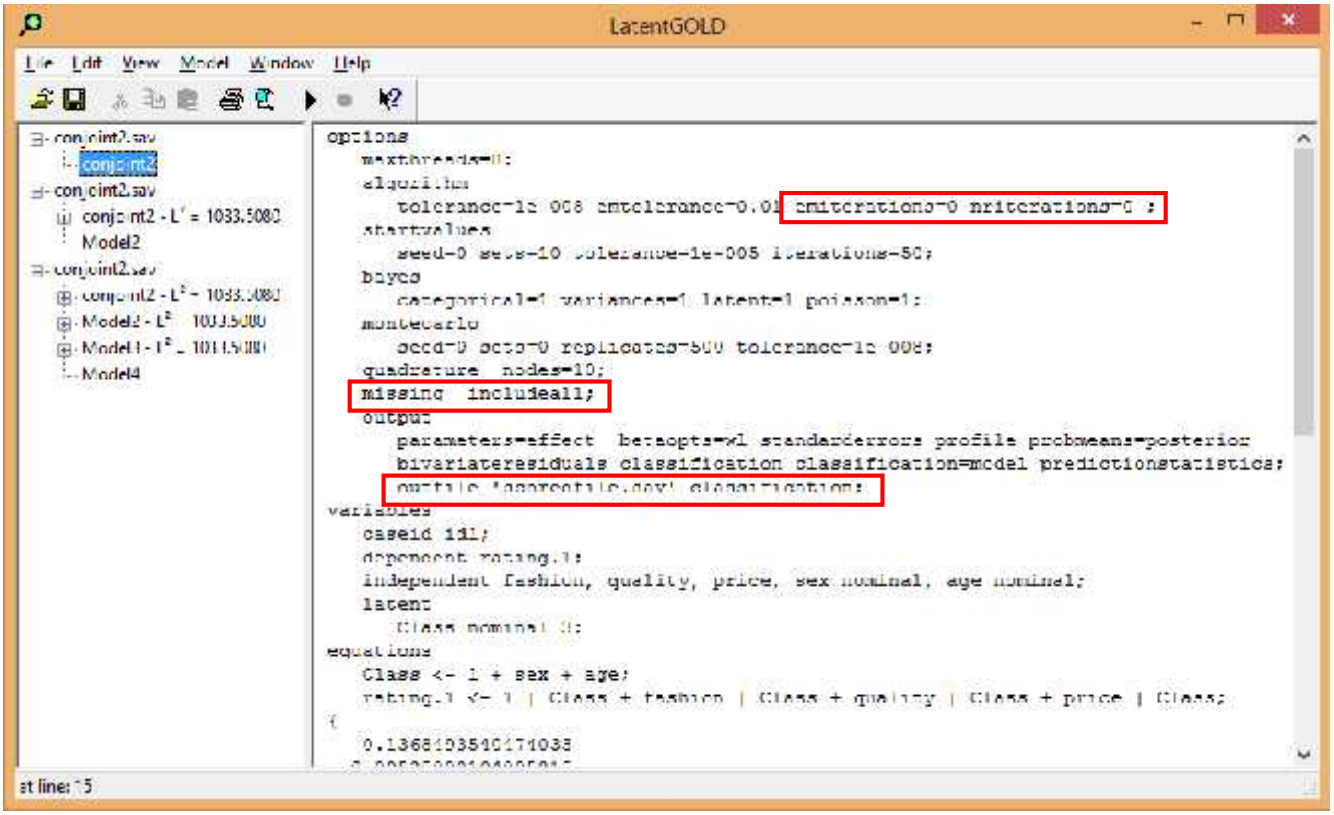
We will now open the model with the saved parameters:

- Click File→Open.
- Select ‘conjoint2.lgs’ and click ‘Open’.

In the syntax editor, make the following changes to the syntax:

- In the algorithm sub-section, change ‘emiterations=250 nriterations=50’ to ‘emiterations=0 nriterations=0’.
- In the Options section, change ‘missing excludeall;’ to ‘missing includeall;’.
- In the Output sub-section, insert ‘outfile ‘scoredfile.sav’ classification;’

Figure 9. Syntax model modified to score new cases.



- Right click on ‘conjoint2’ and click ‘Estimate’.

After estimation has completed, confirm that the $L^2 = 1033.5080$.

Upon estimation, a new SPSS file (‘scoredfile.sav’) will be generated. Open this SPSS file and confirm that all records have been scored.