Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

Markov Tutorial #2: Latent GOLD Longitudinal Analysis of Life Satisfaction

Overview

The goal of this tutorial is to show how Latent GOLD 5.0 can be used to estimate time heterogeneous and mixture latent Markov models, especially useful in explaining unusual time trends.

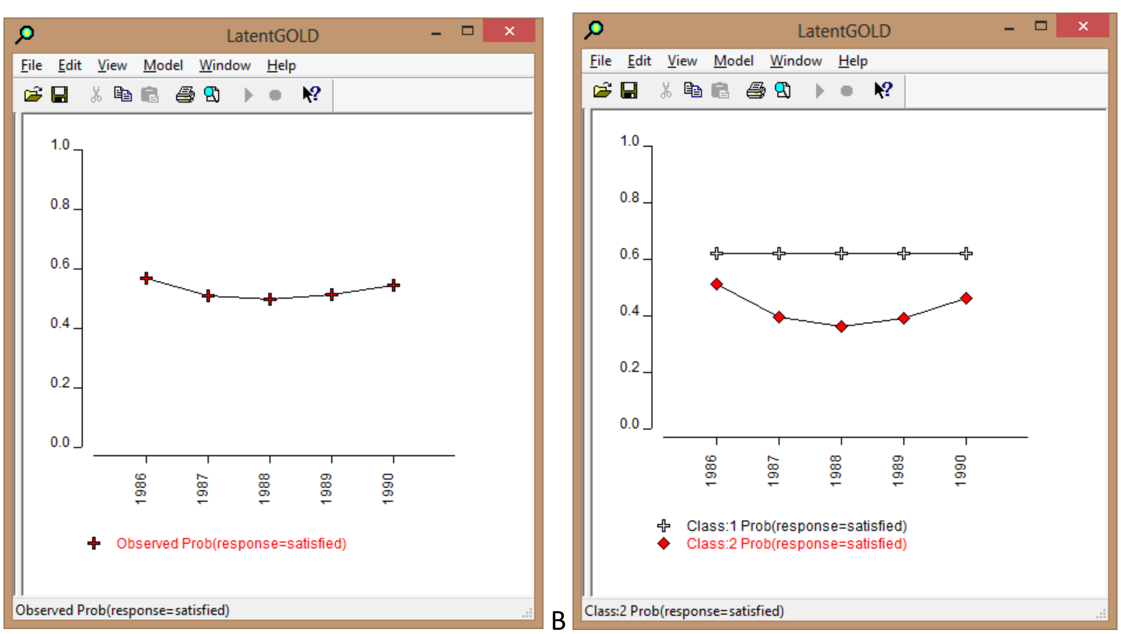
The Data

The data file satisfaction.sav contains a dichotomous indicator of life satisfaction ('response') measured at 5 equidistant time points: Satisfied with life? 1=No or 2=Yes. For further details of these data see:

Langeheine, R. and Van der Pol, F. (2002). "Latent Markov Chains," in Hagenaars, J. and McCutcheon, A. (eds.) Applied Latent Class Analysis, Cambridge University Press, Cambridge, UK

As shown in Figure 1A, the life satisfaction trend in Germany between 1986 and 1990 is nonlinear, the percentage reporting being satisfied decreasing between 1986 and 1988 (time=1-3), and then increasing in 1989 and 1990 (time=4,5). Figure 1B plots the results from a Mover-Stayer latent Markov model showing that the population consists of 2 latent class segments, where the Stayer class (Class 1) shows no change over time.

A



B

Figure 1. (A)Observed sample proportions reporting satisfaction with their lives between 1986 ('time'=1) and 1990 ('time'=5) and (B) predicted class-specific probabilities obtained from a latent Markov model.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

In this tutorial we will show how to obtain the Mover-Stayer latent Markov model (as well as a variety of other models) from these data, where the variable 'ID' is used to distinguish each of the $2^5 = 32$ response patterns and the variable 'weight' is used as a case weight. For example, the *long* format data in Figure 2 shows that 891 respondents expressed dissatisfaction with their life at all 5 time points (i.e., for 'ID=1', 'weight' = 891). In total there are N=5,147 respondents.

(Note: rather than using a case weight, individual case level data with 5,147 records could be used in Latent GOLD.)

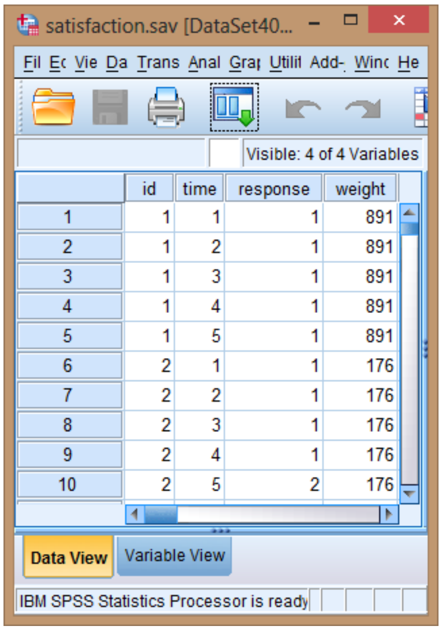


Figure 2. SPSS data file satisfaction.sav where response = 2 corresponds to 'satisfied with life'.

The measurement model component of latent Markov models allows for measurement error in the reported satisfaction responses by estimating the true latent states. Specifically, we will use Latent GOLD's Markov module to estimate a variety of 2-state models, each state corresponding to a response category adjusted for measurement error – a true 'Satisfied' state, and a true 'Not Satisfied' state.

Estimating the Null Model

To estimate the null model of mutual independence between responses at each of the 5 time points:

- From the File menu, open file 'satisfaction.sav'
- Right click on Model 1 and select 'Regression'
- Move 'response' to the Dependent box, right click and select 'nominal'[1].
- Move 'time' to the Predictors box, right click and select 'nominal'.
- Move 'id' to the Case ID box
- Move 'weight' to the Case Weight box

[1] Since response is dichotomous, the model is identical whether treated as nominal or ordinal. Here we apply the nominal scale type in order to provide some simplification in Latent GOLD's output tables.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

- In the Output tab, request the additional output 'EstimatedValues - Regression'

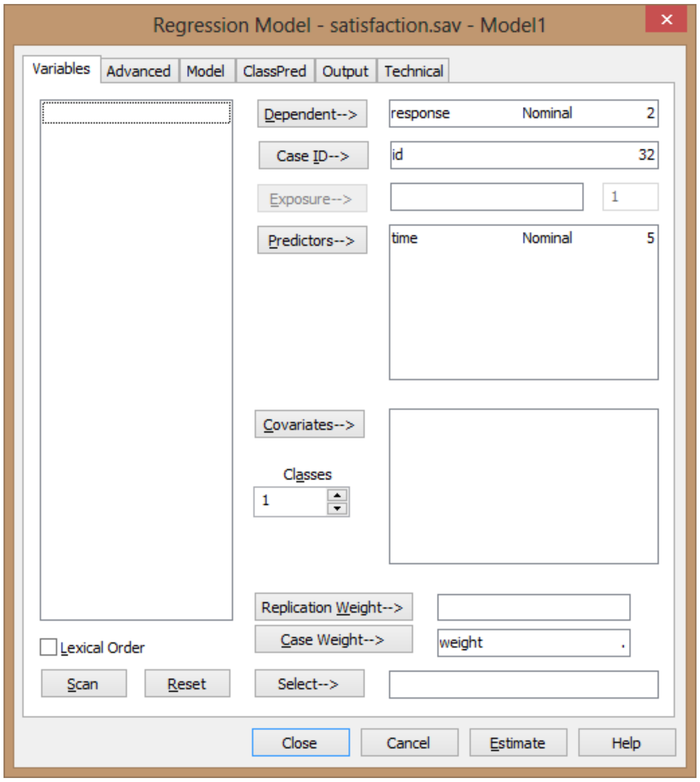


Figure 3. Latent GOLD setup for the null model of mutual independence from the long format data.

- Click Estimate
- For clarity, rename 'Model1' to 'null'.
- Click on the name of the data file 'satisfaction.sav' to display the Model Summary Output
- Then right click in Summary Output pane to retrieve the Model Summary Display, and use it to remove LL, BIC(LL) and to add BIC(L^2).

As indicated in Figure 4, this model fits very poorly and should be rejected (p=5.2E-1060).

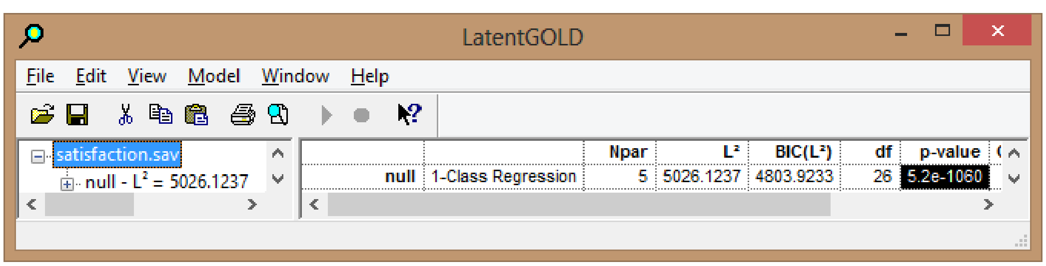


Figure 4. Model Summary Output

The 'EstimatedValues' output contains the predictions for response at each time point. Since the null model corresponds to a 1-class model, the predictions (shown in column 'Class1') match those in the column 'Overall', which are also seen to match the corresponding Observed proportions. The predictions associated with response=2 (satisfied with life) were plotted in Figure 1.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

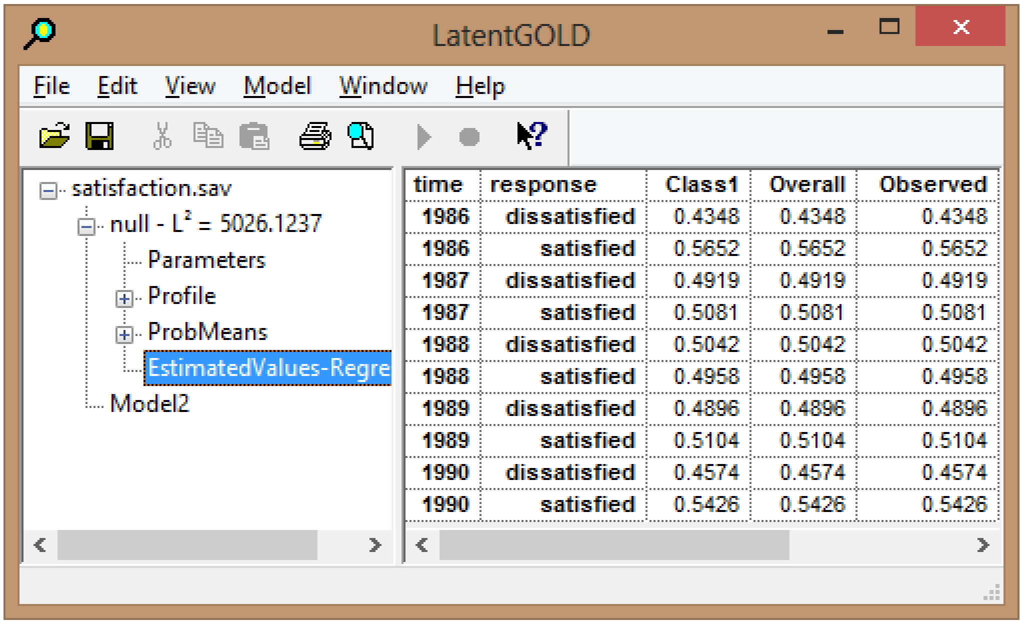


Figure 5. EstimatedValues output for the null model

In longitudinal tutorial #1, we found that a simple 2-state latent Markov model with time independent transition probabilities provided an excellent fit to the data. With the current data, to obtain an adequate fit we will need to make the transition probabilities time dependent.

Estimating a 2-state time-heterogeneous latent Markov Model

The path diagram corresponding to the *time-heterogeneous* latent Markov model is given in Figure 6. This differs from the diagram for the *time-homogeneous* model in that separate transition probability parameters exist for each pair of adjacent time points ($b_1 – b_4$). Compare this with Fig. 14 from longitudinal tutorial #1.

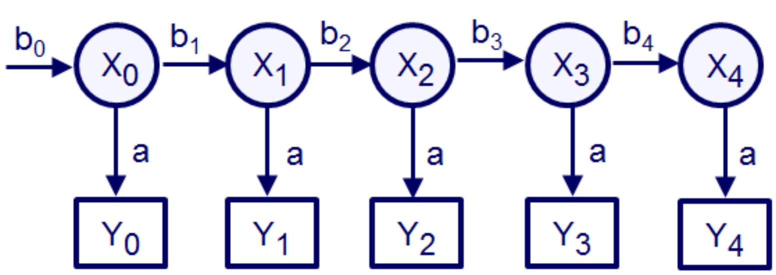


Figure 6. Path diagram for time-heterogeneous latent Markov model, where X_0 and Y_0 refer to the latent state and response variable associated with the initial time point

To estimate the 2-state time-heterogeneous latent Markov model:

- Right click on ‘Model2’ and select ‘Markov’
- Move ‘response’ to the Indicators box.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

- Move ‘time’ to the Covariates box.
- Request 2 states

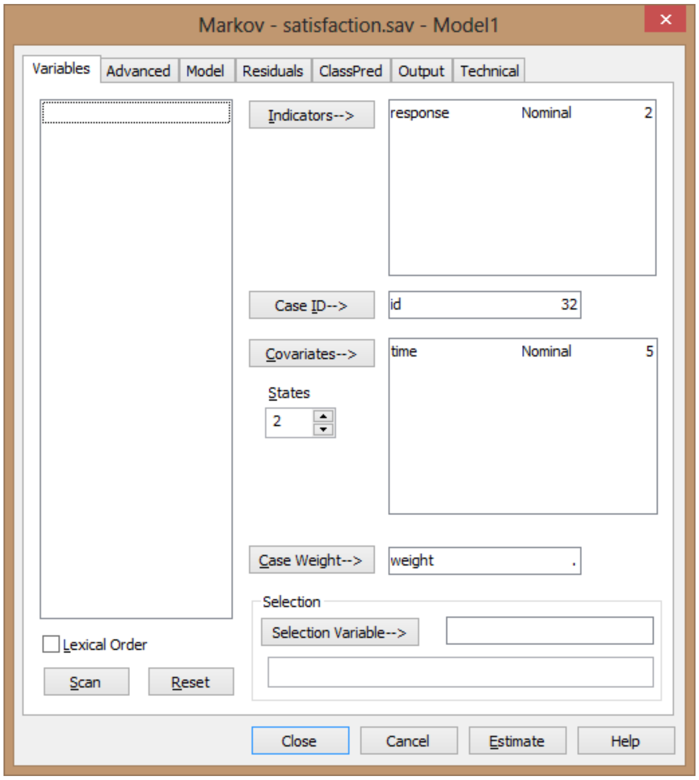


Figure 7. Latent GOLD setup for the 2-state time heterogeneous latent Markov model

- Click Estimate
- Rename ‘Model2’ to ‘2-state time heterogeneous latent Markov’
- Click on the name of the data file ‘satisfaction.sav’ to display the Model Summary Output

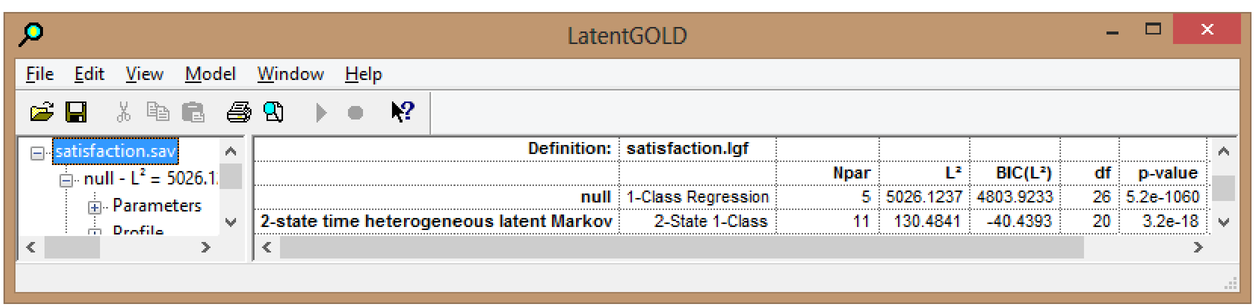


Figure 8. Model Summary Output

Note that the p-value for this model is still less than .05, so overall this model with Npar=11 parameters still does *not* provide an adequate fit to the data. However, longitudinal bivariate residuals (L-BVRs) show that this model *does* reproduce the observed time trend and also explains the 1st and 2nd order

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

autocorrelations adequately[2]. To display these BVRs to see that they are all non-significant (less than 3.84):

➢ Click on bivariate residuals output for the Markov model

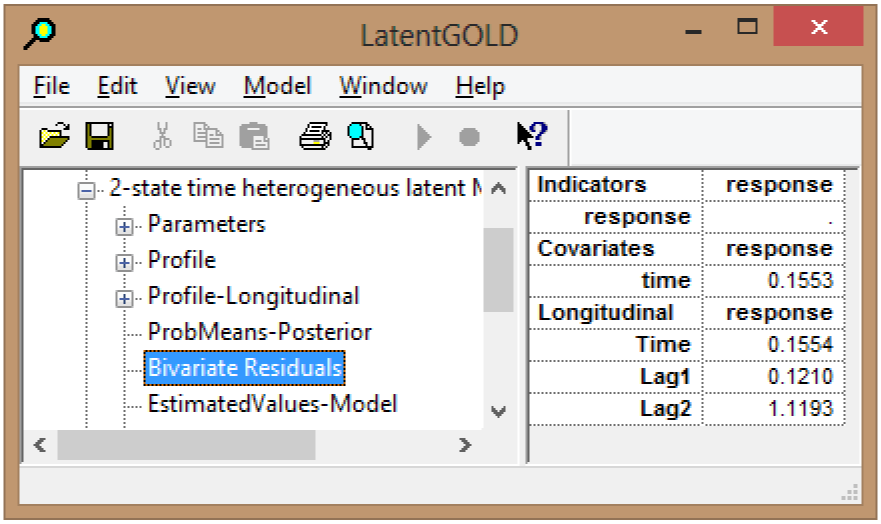


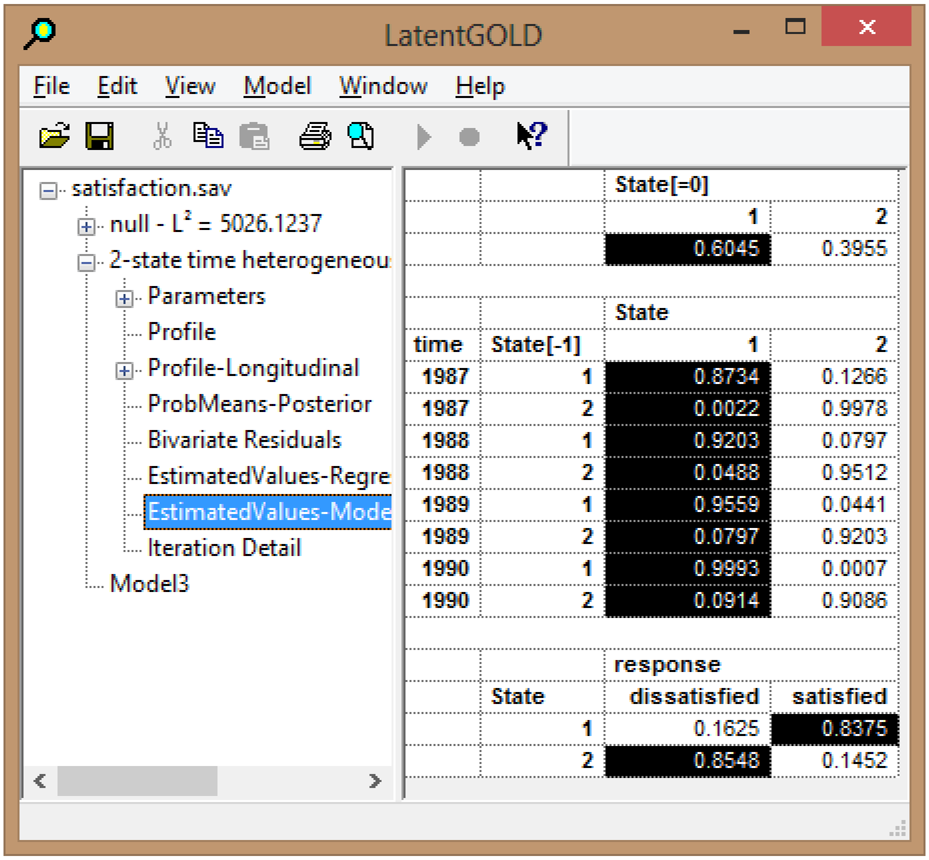
Figure 9. Longitudinal Bivariate Residuals for the 2-state time heterogeneous LM model

Note: These Lag1 and Lag2 L-BVRs represent a substantial reduction in the amount of 1st and 2nd order autocorrelation in the observed data, which could be quantified by the Lag1 and Lag2 L-BVRs based on the null model – Lag1(null) = 3673.06 and Lag2(null) = 2222.34. (While the L-BVRs are not available as output from the null model when estimated from the GUI, they can be requested as output from the syntax module.)

Click on 'EstimatedValues - Model' to display the 11 distinct probability parameter estimates:

[2] See the Appendix of longitudinal data analysis Tutorial 1 for an introduction to the longitudinal BVRs. These were obtained for this null model using the Latent GOLD syntax module.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)



		State[=0]	
		1	2
		0.6045	0.3955

		State	
time	State[-1]	1	2
1987	1	0.8734	0.1266
1987	2	0.0022	0.9978
1988	1	0.9203	0.0797
1988	2	0.0488	0.9512
1989	1	0.9559	0.0441
1989	2	0.0797	0.9203
1990	1	0.9993	0.0007
1990	2	0.0914	0.9086

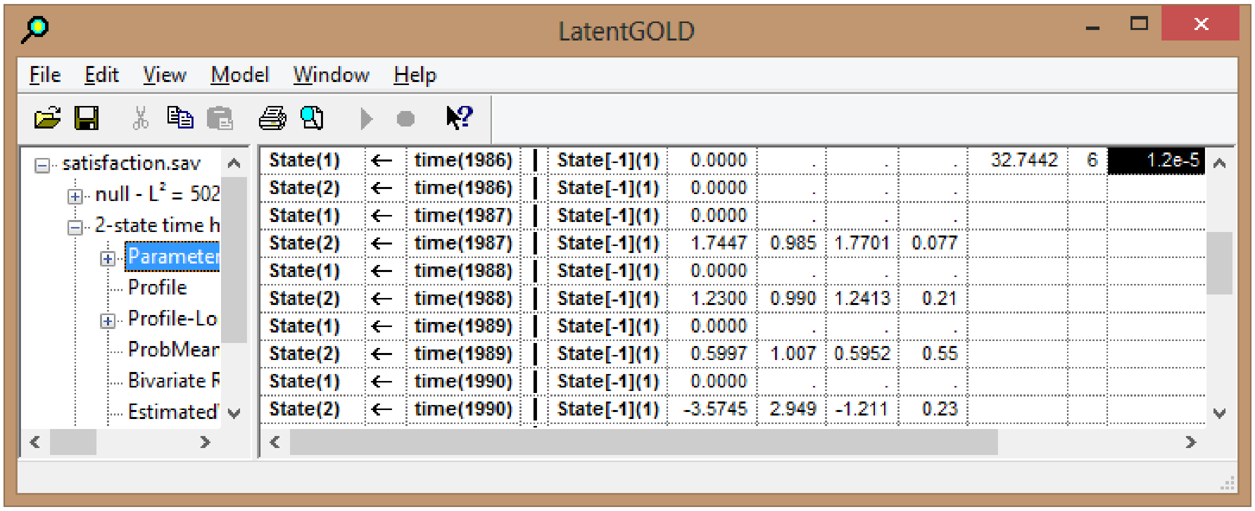
		response	
	State	dissatisfied	satisfied
	1	0.1625	0.8375
	2	0.8548	0.1452

Figure 10. Estimated values for 2-state LM model

- **The 2 distinct measurement model probability parameters** (highlighted at the bottom) shows that a[1] = 83.75% of the time respondents in state 1 report that they are satisfied with their life ('response' = 2), which reflects measurement error of 100% - 83.75% = 16.25%. Thus, State 1 can be named 'Satisfied'. Similar reasoning allows us to name State 2 'Dissatisfied', a similar measurement error (14.52%) being estimated for those saying they are dissatisfied.
- **The 1 distinct initial state probability parameter** at the top shows that b_0 = 60.45% are in State 1 (Satisfied) in 1986, the initial year in this analysis sample.
- **The 8 transition probabilities** show that over time 'Satisfied' respondents are more likely to remain in the Satisfied state the next year. For example, the probability of remaining Satisfied in 1987 (time=2) is b1[1] = .8734, while the probability of Satisfied persons in 1989 remaining Satisfied in 1990 is b1[2] = .9993.

From the Parameters output (Figure 11) it can be seen that the differences in the transition probabilities over time are statistically significant (p=.000012).

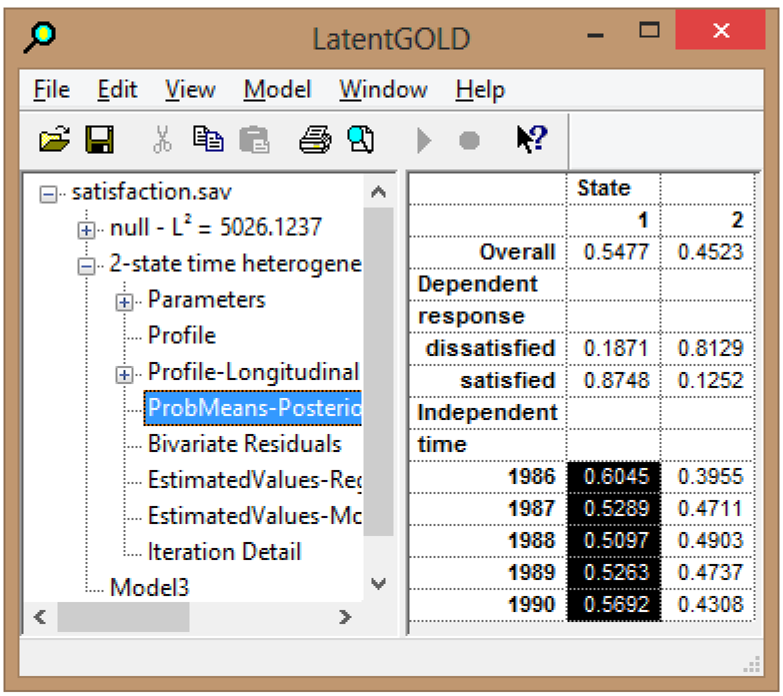
Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)



State(1)	←	time(1986)	State[-1](1)	0.0000	.	.	.	32.7442	6	1.2e-5
State(2)	←	time(1986)	State[-1](1)	0.0000	.	.	.			
State(1)	←	time(1987)	State[-1](1)	0.0000	.	.	.			
State(2)	←	time(1987)	State[-1](1)	1.7447	0.985	1.7701	0.077			
State(1)	←	time(1988)	State[-1](1)	0.0000	.	.	.			
State(2)	←	time(1988)	State[-1](1)	1.2300	0.990	1.2413	0.21			
State(1)	←	time(1989)	State[-1](1)	0.0000	.	.	.			
State(2)	←	time(1989)	State[-1](1)	0.5997	1.007	0.5952	0.55			
State(1)	←	time(1990)	State[-1](1)	0.0000	.	.	.			
State(2)	←	time(1990)	State[-1](1)	-3.5745	2.949	-1.211	0.23			

Figure 11. Parameters output for the time heterogeneous latent Markov model showing that the differences in the transition probabilities over time are statistically significant

- ➢ Click on ProbMeans to see that the trend in the satisfaction state (State 1) corresponds closely to the observed trend plotted in Figure 1.



	State	
	1	2
Overall	0.5477	0.4523
Dependent		
response		
dissatisfied	0.1871	0.8129
satisfied	0.8748	0.1252
Independent		
time		
1986	0.6045	0.3955
1987	0.5289	0.4711
1988	0.5097	0.4903
1989	0.5263	0.4737
1990	0.5692	0.4308

Figure 12. ProbMeans-Posterior for 2-state LM model

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

Mixture 2-state time-heterogeneous latent Markov with 2 latent classes

Next, we will see that including 2 classes where the classes differ with respect to 1) their initial state probabilities and 2) their transition probabilities yields an adequate fit to these data.

- Double click on the 2-state time-heterogeneous latent Markov model
- In the Advanced tab, request 2 classes

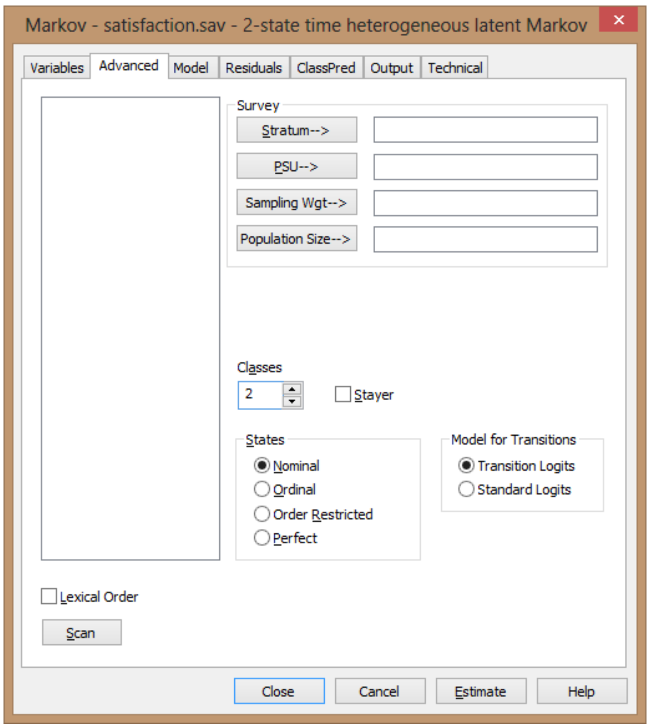
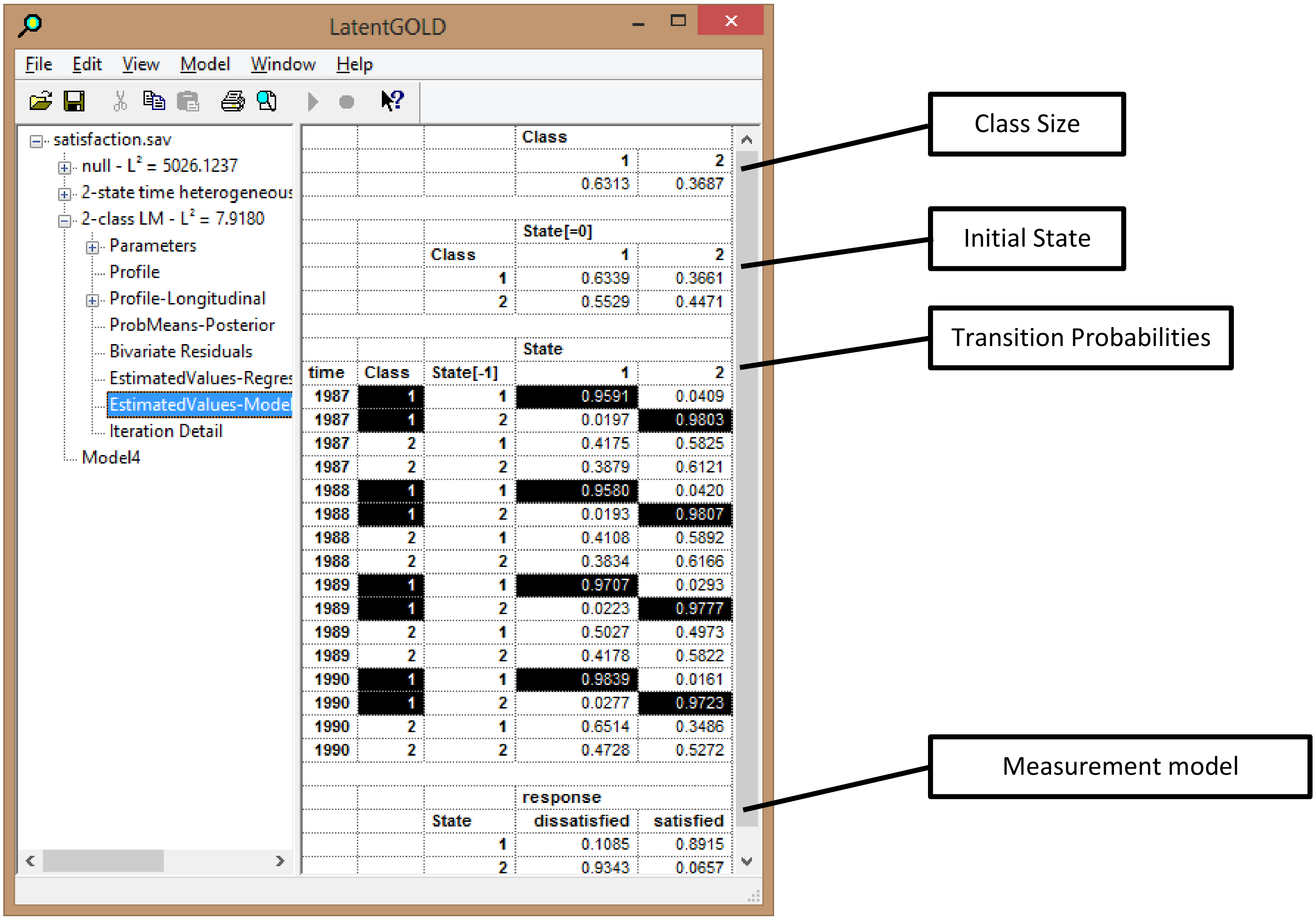


Figure 13. Requesting 2 classes in the Advanced tab for Markov model

- Click Estimate
- Rename ‘Model4’ to ‘2-class LM’
- Click on the EstimatedValues-Model output

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)



	Class	
	1	2
	0.6313	0.3687

	State[=0]	
Class	1	2
1	0.6339	0.3661
2	0.5529	0.4471

time	Class	State[-1]	State 1	State 2
1987	1	1	0.9591	0.0409
1987	1	2	0.0197	0.9803
1987	2	1	0.4175	0.5825
1987	2	2	0.3879	0.6121
1988	1	1	0.9580	0.0420
1988	1	2	0.0193	0.9807
1988	2	1	0.4108	0.5892
1988	2	2	0.3834	0.6166
1989	1	1	0.9707	0.0293
1989	1	2	0.0223	0.9777
1989	2	1	0.5027	0.4973
1989	2	2	0.4178	0.5822
1990	1	1	0.9839	0.0161
1990	1	2	0.0277	0.9723
1990	2	1	0.6514	0.3486
1990	2	2	0.4728	0.5272

	response	
State	dissatisfied	satisfied
1	0.1085	0.8915
2	0.9343	0.0657

Figure 14. EstimatedValues for mixture 2-state time-heterogeneous LM model with 2 latent classes

The highlighted cells in Figure 14 show that for class 1 the probability of staying in the same state is close to 1. To test the hypothesis that these small differences from 1 are due to chance, we will estimate a Mover-Stayer model which restrict these probabilities to 1, making this a 'Stayer' class.

Estimating the Mover-Stayer time-heterogeneous latent Markov model

- Double click on the '2-class LM' model
- In the Advanced tab, check the 'Stayer' box to make one of the 2 classes a 'Stayer' class

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

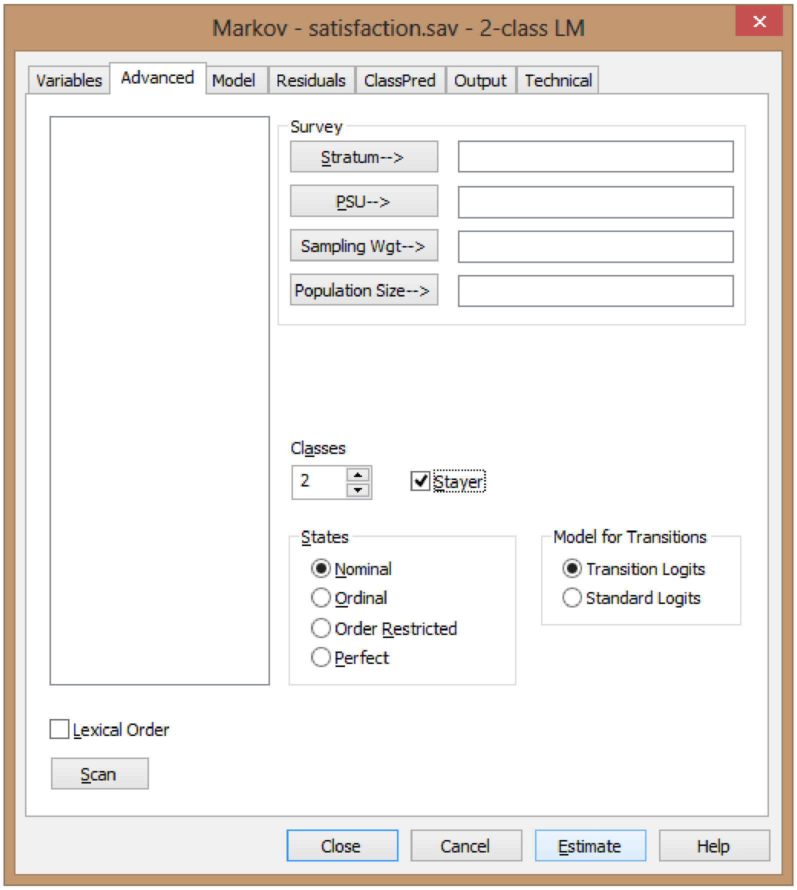


Figure 15. Selecting 'Stayer' in the Advanced tab for Markov model

- ➢ Click Estimate
- ➢ Rename 'Model5' to 'Mover-Stayer LM'
- ➢ Click on the data file name to display the Model Summary output

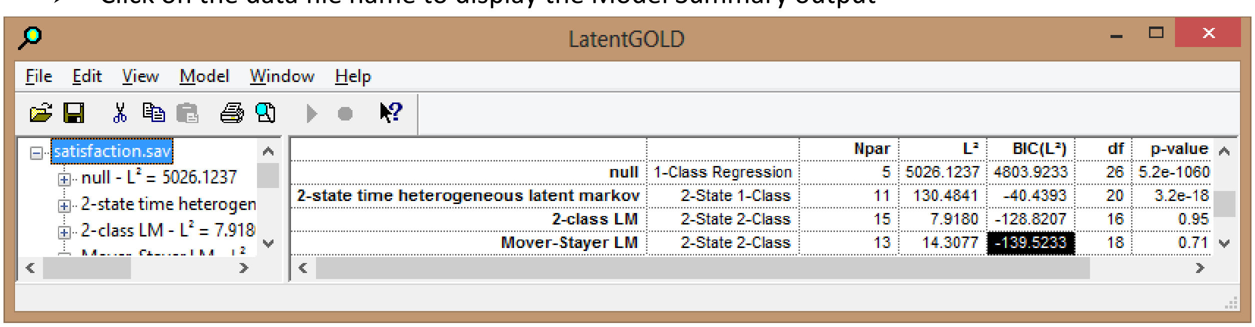


Figure 16. Model Summary Output

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

We see that both the unrestricted and restricted 2-class models provide an adequate fit to the data (p=.95 and .71 respectively), the BIC statistic preferring the restricted (Mover-Stayer) model. To verify the transition probabilities for class 1 (the Stayer class) now correspond to the identity matrix,

- ➢ Click on 'EstimatedValues – Model' output for the Mover-Stayer LM model

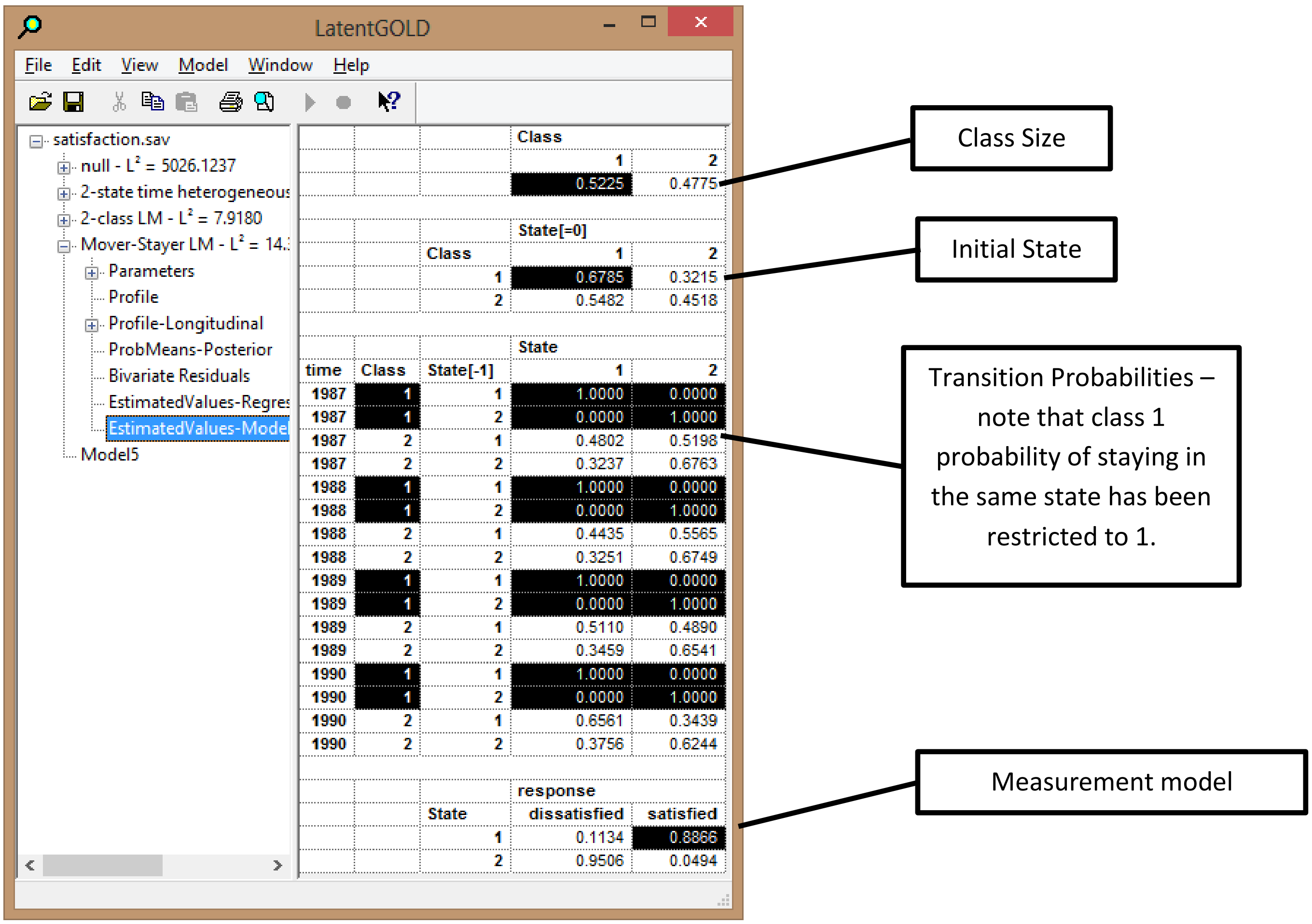


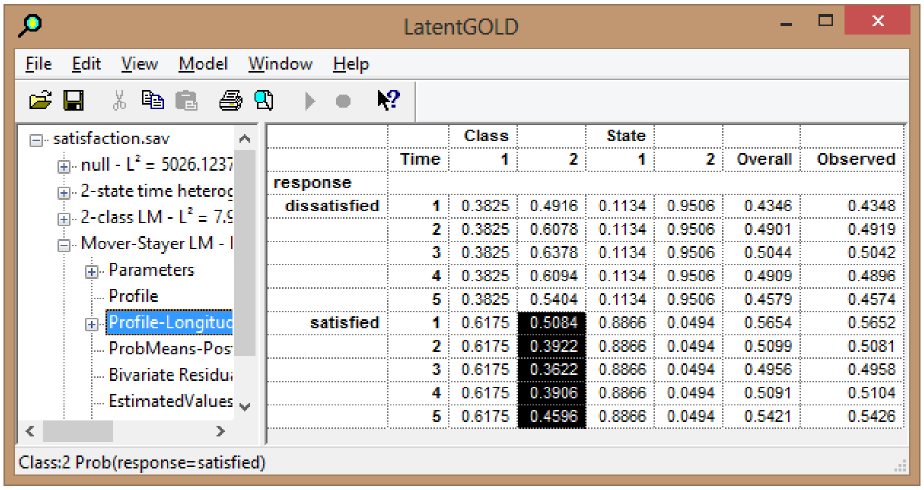
Figure 17. 'EstimatedValues–Model' output showing Class 1 as a Stayer class.

The Estimated Values output (Figure 17) shows that 52.25% of respondents are in the Stayer class, who tend to be mostly Satisfied with their lives throughout this 5 year period -- 67.85% are in state 1 ('Satisfied' state) initially and remain in that state. In contrast, among respondents whose life satisfaction *changed* during this 5 year period (the 'Mover' class), fewer (54.82%) were in the Satisfied state during the initial year.

Also, as shown at the bottom of this output, there is less measurement error among those saying they are dissatisfied (4.94% error) than those who express 'Satisfaction' (11.34% error). To see how the 2 classes differ with respect to their response trends,

- ➢ Click on Profile-Longitudinal

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)



		Class		State			
	Time	1	2	1	2	Overall	Observed
response							
dissatisfied	1	0.3825	0.4916	0.1134	0.9506	0.4346	0.4348
	2	0.3825	0.6078	0.1134	0.9506	0.4901	0.4919
	3	0.3825	0.6378	0.1134	0.9506	0.5044	0.5042
	4	0.3825	0.6094	0.1134	0.9506	0.4909	0.4896
	5	0.3825	0.5404	0.1134	0.9506	0.4579	0.4574
satisfied	1	0.6175	0.5084	0.8866	0.0494	0.5654	0.5652
	2	0.6175	0.3922	0.8866	0.0494	0.5099	0.5081
	3	0.6175	0.3622	0.8866	0.0494	0.4956	0.4958
	4	0.6175	0.3906	0.8866	0.0494	0.5091	0.5104
	5	0.6175	0.4596	0.8866	0.0494	0.5421	0.5426

Figure 18. Longitudinal-Profile output for Mover-Stayer LM model.

The Longitudinal-Profile output shows that among Stayers (Class=1), 61.75% report being satisfied each year. Among respondents whose life satisfaction changed during this 5 year period (Class 2 = ‘Movers’), 50.84% reported being satisfied with their lives during the initial year, falling to the 36% - 39% range over the next 3 years before increasing to about 46% during the 5th year.

The final 2 columns show how very closely the *predicted* response trend (Overall column) agrees with the corresponding *observed* proportions (Observed column) which were plotted earlier (recall Figure 1).

To view the longitudinal profile plot

- ➢ Click on the “+” next to Profile Longitudinal
- ➢ Click on Longitudinal-Plot
- ➢ Right click on the plot to open the ‘Longitudinal-Plot Display’

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

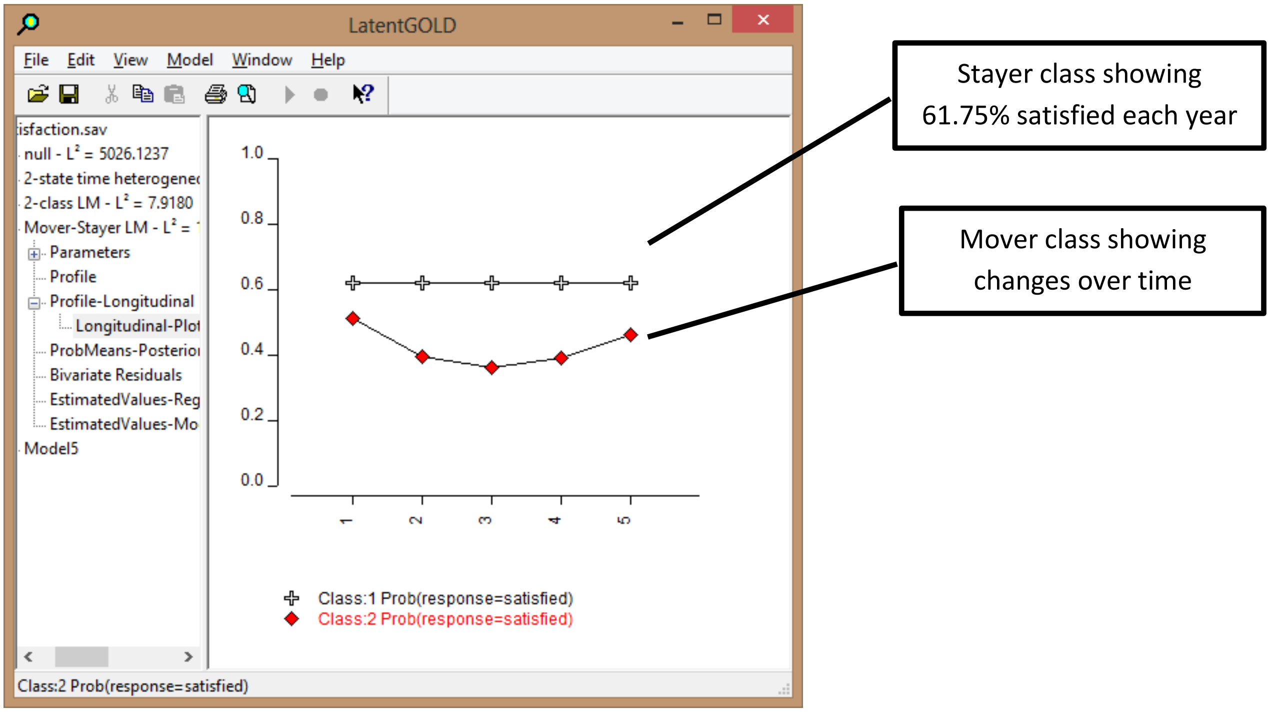


Figure 19. Longitudinal Profile Plot for the Mover-Stayer LM model, with the Mover class (Class 2) highlighted.

As reported above (Figure 17), 52.25% are in the Stayer class, the remaining respondents in the Mover class. These class sizes can be used as weights to obtain the predicted probability of being satisfied in each year. These predicted probabilities are reported in the Overall column of the Profile-Longitudinal output (recall Figure 18). To add these predictions to the plot,

- Right click on the plot to open the Longitudinal-Plot Display
- In the ‘Classes’ box, scroll down and check the ‘Overall’ box

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

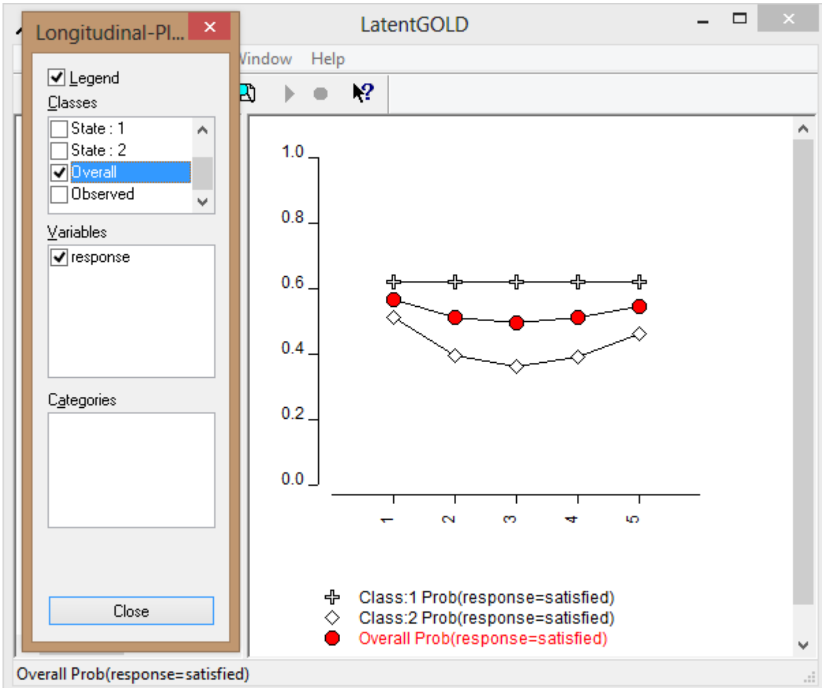


Figure 20. Longitudinal-Plot with predicted probability of being satisfied ('Overall Prob') appended.

As shown in the Longitudinal-Profile output (Figure 18), the predicted probabilities match the observed proportions to 2 decimal places. This close agreement can be shown graphically as follows:

- Check 'Observed' to append the observed proportions to the plot
- Uncheck 'Class: 1' and 'Class: 2' to remove the class specific response probabilities

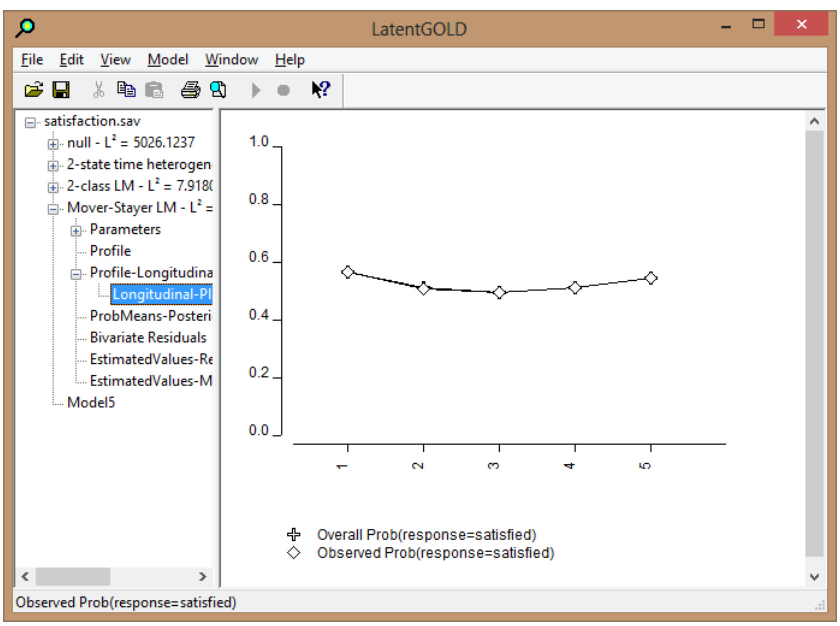


Figure 21. Longitudinal plot showing overlap between predicted probability and observed proportion.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

This close agreement is quantified by the 'Time' longitudinal BVR, which is very close to zero (0.0281). To view this as well as the small (non-significant) Lag1 and Lag2 longitudinal BVRs under this Mover-Stayer LM model,

➢ Click on the Bivariate Residuals

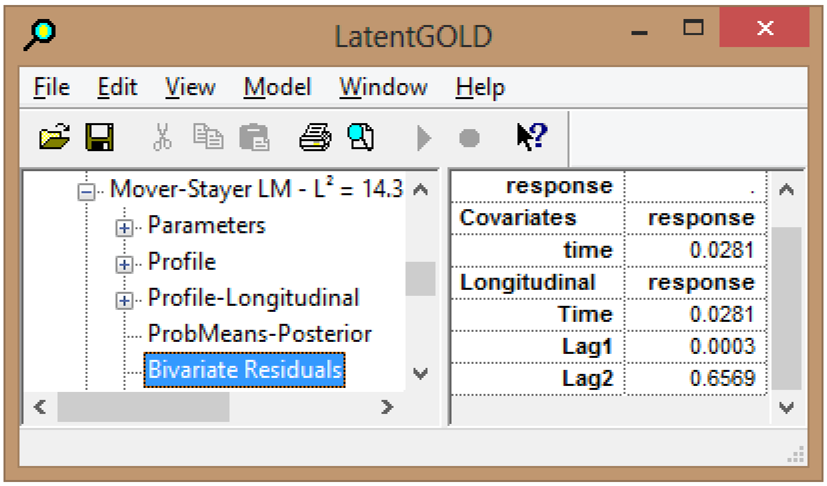


Figure 22. L-BVRs for the Mover-Stayer model.

These small residuals support the excellent fit provided by this Mover-Stayer time-heterogeneous LM model.

Viewing the LG-Syntax Equations

To see the equations corresponding to each of the models created in this tutorial,

➢ Right click on the 'satisfaction.sav', and select 'Generate Syntax'.

A separate (new) syntax tree will appear above the GUI tree as shown in Figure 23. Note that the GUI model names are preserved in the syntax.

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

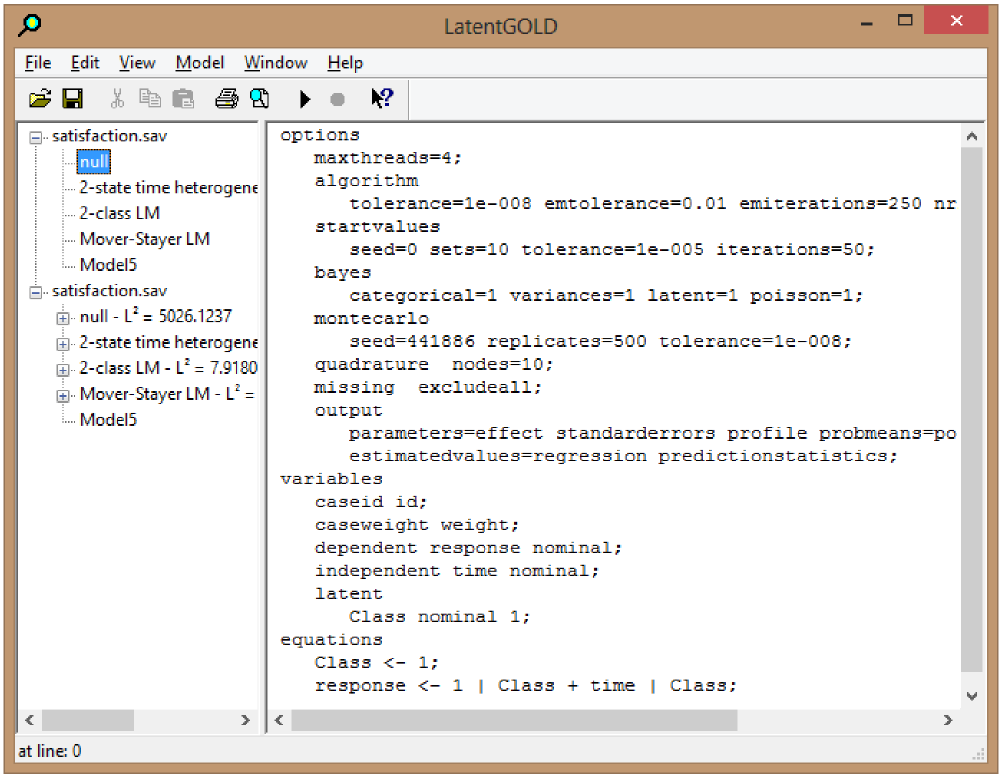


Figure 23. New syntax tree above the GUI tree.

➢ Click on the '2-class LM' model

Note the equations syntax for this model (annotated in Figure 24). Here, we allow different state probabilities for each class. Recall the 'EstimatedValues-Model' output shown in Figure 14 for the 2-class LM model, showing for class 1 the probability of staying in the same state is close to 1.

➢ Right click on the '2-class LM' model and select 'Estimate'

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

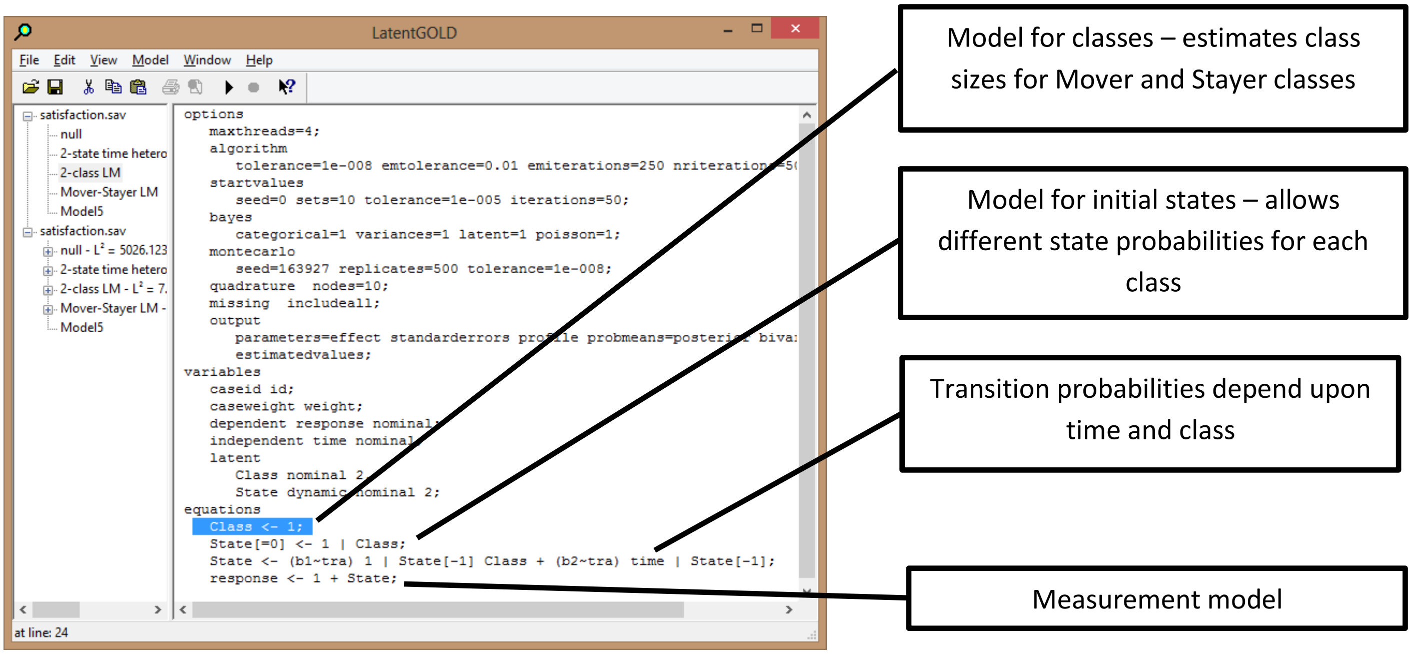


Figure 24. Syntax model specification for estimating 2-class LM model.

- ➢ Click on the 'Mover-Stayer LM'model

Note the equations syntax for this model (annotated in Figure 25). You can see the syntax used to restrict these probabilities to 1, making this a 'Stayer' class (recall Figure 17).

- ➢ Right click on the 'Mover-Stayer LM' model and select 'Estimate'

Copyright © 2013 by Statistical Innovations Inc. All rights reserved. (version: 5/10/2013)

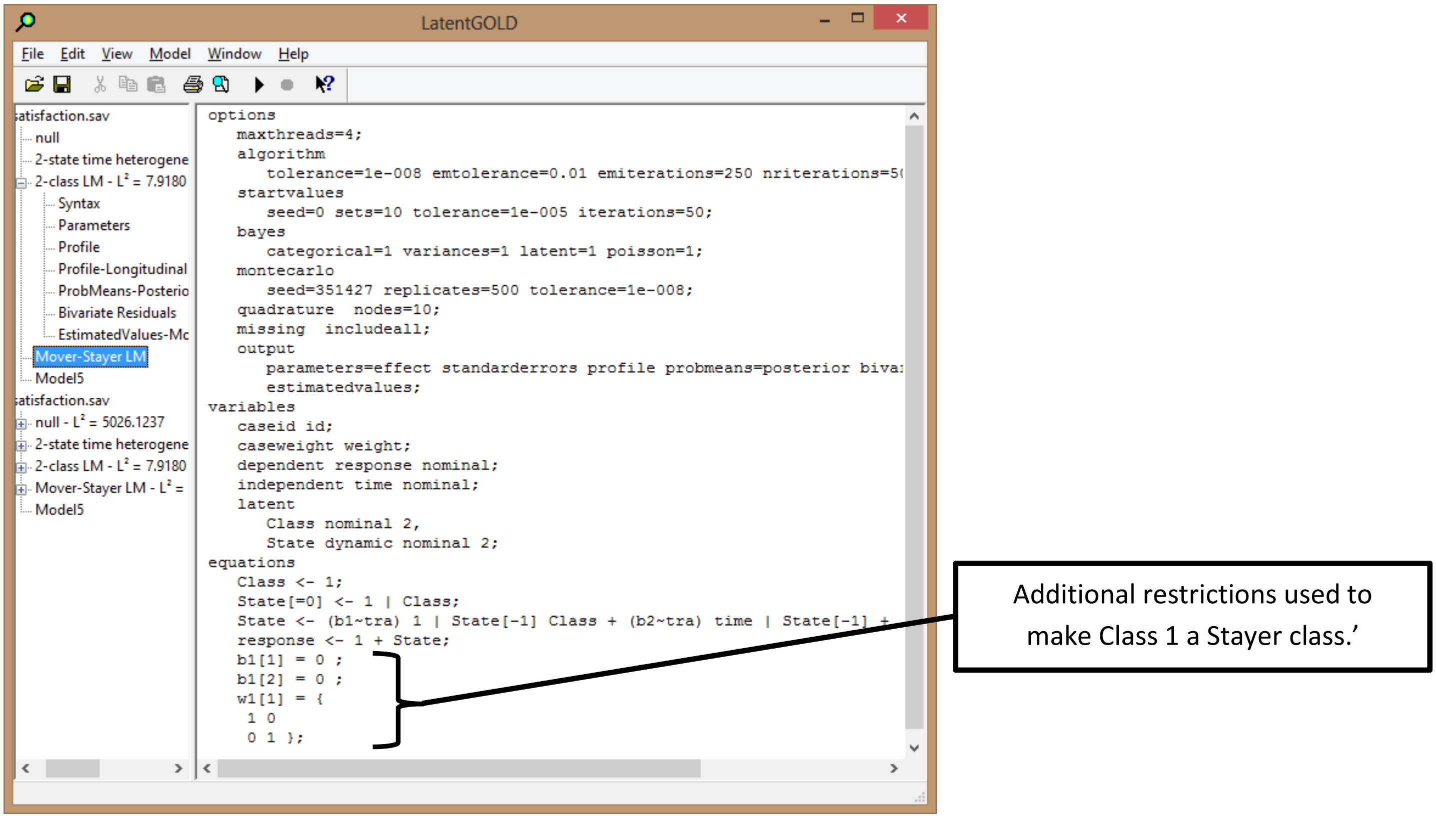


Figure 25. Syntax model specification for estimating Mover-Stayer model.

Summary

The satisfaction data used in this tutorial shows a decline followed by an increase in satisfaction with one's life during the 5 year period analyzed. These data required a time-heterogeneous mixture LM model consisting of 1 Mover and 1 Stayer class to provide an adequate fit to the data. If additional data were available on the respondents, the model could be extended further to include covariates in order to determine how the Movers and Stayers differ in terms of the available characteristics.

Longitudinal tutorials 1 and 2 were limited to include only 5 time points, 1 dichotomous response variable and 2 latent states. More general models can be estimated with the Latent GOLD 5.0 GUI to analyze data containing hundreds of time points, and multiple response variables of varying scale types (nominal, ordinal, count, continuous). In addition, latent Markov models can also be estimated with only a single case, such models typically known as hidden Markov models.

Longitudinal analysis Tutorial 3 provides an additional example of using Latent GOLD 5.0 GUI to estimate a latent Markov model, this example consisting of sparse panel data were obtained at unequal time intervals across 9 panel waves with time constant and time varying predictors where waves of the panel survey. In addition, the syntax version of Latent GOLD 5.0 provides options for estimating many more general models.