Tutorial #7A: Latent Class Growth Model (# seizures)

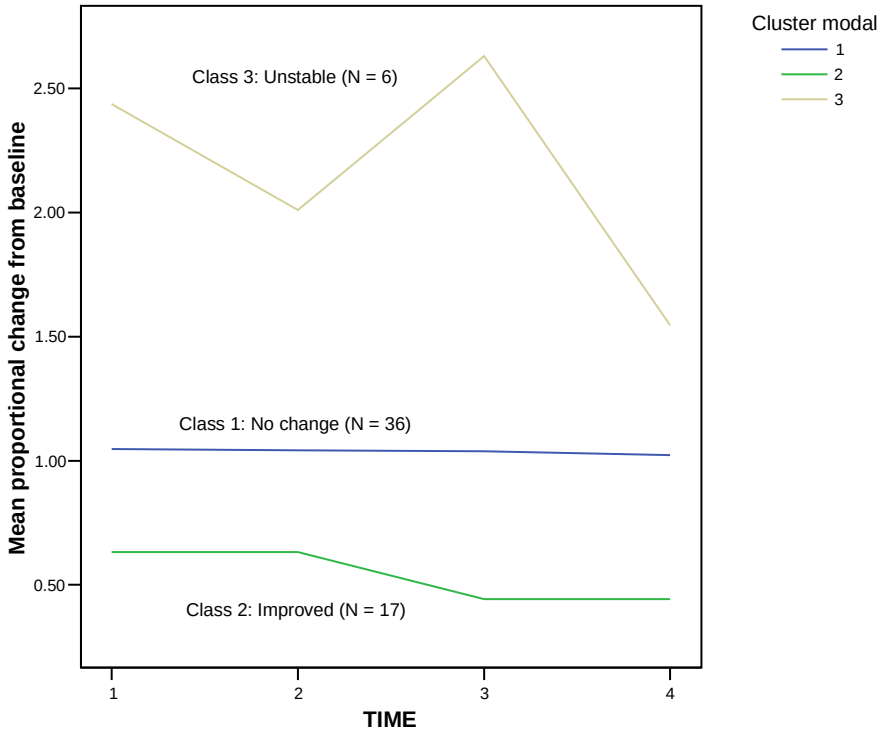


Figure 1. A reanalysis of a longitudinal data set on counts of epileptic seizures using latent class growth models. (Data source: Thall and Vail, 1990)

What you will learn:

- To use Latent GOLD to identify distinct latent class growth trajectories in the data.

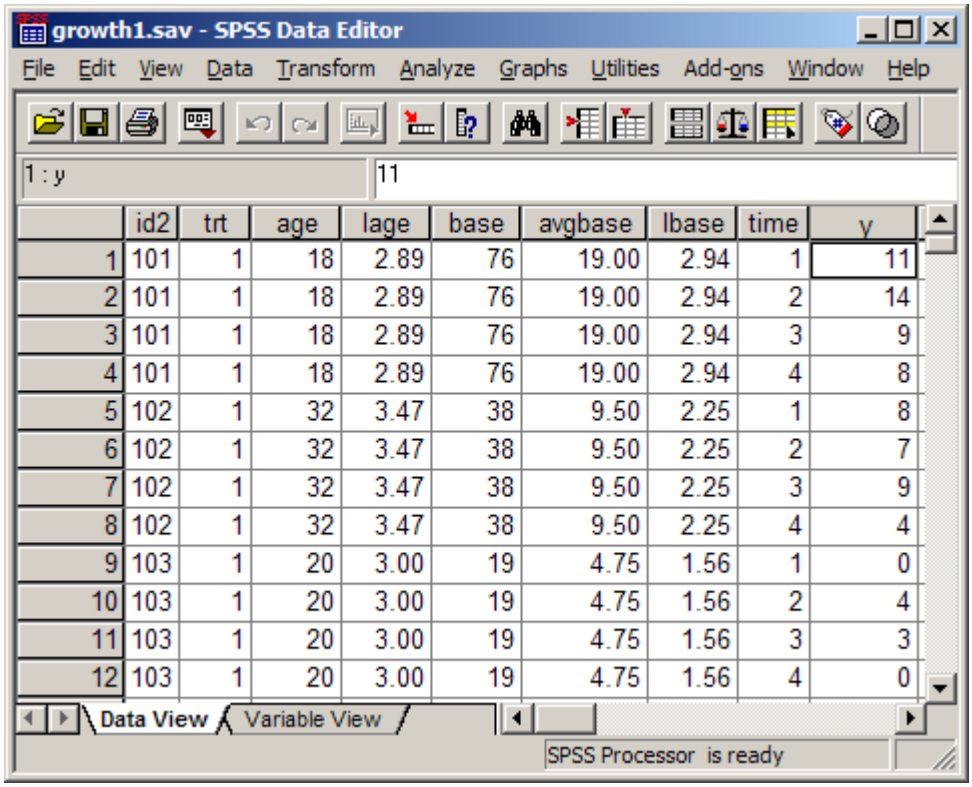
- To name the identified latent class subgroups based on their growth patterns
  - classify 36 cases as unchanged (Class 1 above)
  - classify 17 cases as improved (Class 2 above)
  - classify the remaining 6 cases as Unstable (Class 3 above)

- How to estimate a latent class growth model (Poisson mixture model) to these data that shows those receiving the drug treatment were significantly more likely than the placebo group to improve and significantly less likely to show no change over their baseline seizure rate ($p = .02$).

The Data

In this study, 59 epileptics were randomly assigned to either an anti-seizure medication Progabide (TRT = 1; n=31) or a placebo (TRT = 0; n=28) as an adjunct to other chemotherapy. For each, a base number of seizures was recorded for 8 weeks before the drugs were administered (BASE) and then for four consecutive 2 week periods (TIME) each preceding a visit to the clinic. We also know the age (AGE) of each participant.

Data for the first 3 cases are shown below:



	id2	trt	age	lage	base	avgbase	lbase	time	y
1	101	1	18	2.89	76	19.00	2.94	1	11
2	101	1	18	2.89	76	19.00	2.94	2	14
3	101	1	18	2.89	76	19.00	2.94	3	9
4	101	1	18	2.89	76	19.00	2.94	4	8
5	102	1	32	3.47	38	9.50	2.25	1	8
6	102	1	32	3.47	38	9.50	2.25	2	7
7	102	1	32	3.47	38	9.50	2.25	3	9
8	102	1	32	3.47	38	9.50	2.25	4	4
9	103	1	20	3.00	19	4.75	1.56	1	0
10	103	1	20	3.00	19	4.75	1.56	2	4
11	103	1	20	3.00	19	4.75	1.56	3	3
12	103	1	20	3.00	19	4.75	1.56	4	0

Figure 2. Data for the first 3 cases

The Poisson Mixture Model

Let Y_{ixt} denote the number of epileptic fits during the 2 weeks prior to visit T=t for case i. Case i is assumed to belong to one of K unobservable latent classes x=1,2,…,K. We assume that Y_{ixt} follows the Poisson mixture model given below:

$$\ln(Y_{ixt}) = \beta_0 \ln(Y_{0i}) + \alpha_{.x} + \beta^T_{t.x} + \varepsilon_{ixt} \qquad t = 1,2,3,4$$

where Y_{0i} = base rate of seizures for case i per 2-week baseline period
= BASE /4 (labeled AVGBASE in the .sav file shown above)

$\alpha_{.x}$ is the random intercept measuring the overall change from the baseline

and $\beta_{t.x}^{T}$ is the random effect associated with the t[th] 2-week treatment period
Hence, each latent class x identifies a distinct pattern of change in the seizure rate from the baseline to the treatment period.

For identification of the $\beta_{t.x}^{T}$, we use effect coding:

$$\sum_{t=1}^{4} \beta_{t.x}^{T} = 0$$

In our final model, β_0 will be modeled as a class independent offset – i.e., β_0 =1, which implies that the expected % change in the seizure rate from the baseline to period t is:

$$E(Y_{ixt} / Y_{0i}) = \exp(\alpha_{.x} + \beta_{t.x}^{T})$$

Thus, the growth trajectory associated with latent class x is given by:

$$\exp(\alpha_{.x} + \beta_{t.x}^{T}) \qquad \text{t=1,2,3,4}$$

Tutorials #7A and #7B illustrate somewhat different approaches for using Latent GOLD to estimate 1) the trajectories and 2) the treatment effect. These different approaches agree on the following results:

- Latent class #1 contains approximately 57% of the cases who show basically no change from the baseline rate.
- Class #2 (31%-32% of the cases) show a significant reduction in seizure rate.
- The remaining class(es) consist of 6 unstable cases who show a substantial increase in at least one of the treatment periods. These 6 cases were all identified as outliers in an earlier analysis of these data by Rabe-Hesketh and Skrondal (2004).
- Those treated with Progabide were significantly less likely to be in class 1 ('no change') and significantly more likely to be in class 2 ('improved') than the Placebo group.

In tutorial #7A, our analysis consists of two steps. First, we estimate a *pure* (unsupervised) growth model to identify the K different growth patterns over the four post time periods. This growth model *pure* in the sense that covariate information (AGE and TREATMENT status of the cases) is not taken into account. During step 2 we assess the relationship between these covariates and the different classes.

In tutorial #7B, we estimate the treatment effect and all the model parameters simultaneously (i.e., in a single step) by specifying TREATMENT as an *active* as opposed to *inactive* covariate.

The models and methods illustrated here are simpler to apply than approaches based on Generalized Estimating Equations that have used in conjunction with these data by others (Diggle, et. al, 1994; Lee, 2004). All estimates are maximum likelihood.

Setting up the Model

To retrieve the setups for these models:

- FILE OPEN epil.lgf
- Double click on 'Model1'

The Variables tab appears as follows:

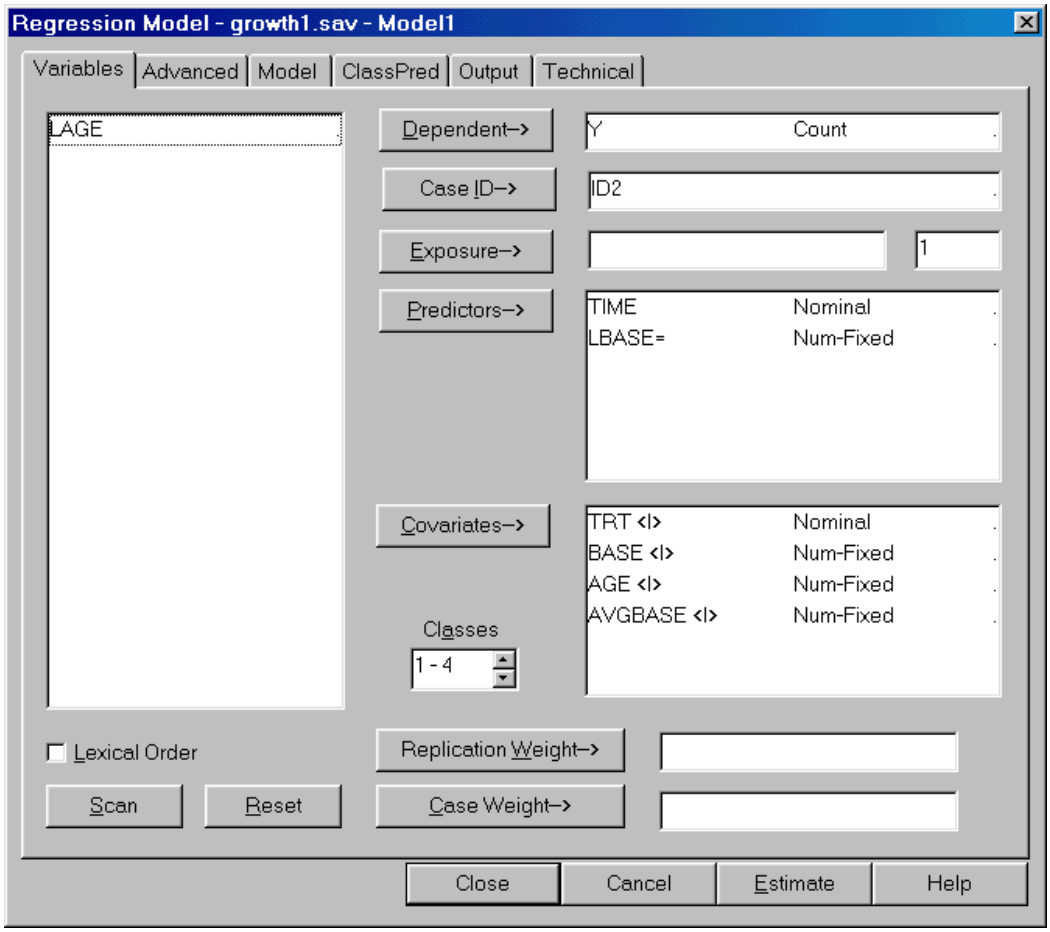


Figure 3. Variables Tab for epil.lgf

The scale type for the dependent variable Y is set to 'Count' which means that a Poisson regression model will be estimated for each latent class.

ID2 is used in the case ID box, indicating that there are multiple records for each case.

The predictors TIME and LBASE are included in the Predictors box:

- TIME is specified as a nominal predictor which allows the estimated time trend to take on any pattern (such as a reduction during period 1 followed by an increase during period 2). Separate distinct time trends are identified for each class (i.e., random effects are estimated for TIME).

- The predictor LBASE is treated as class independent which means that this estimate is restricted to be equal in each class. This is indicated by the symbol '=' which appears next to the variable LBASE in the Predictors box. Thus, a single *fixed effect* will be estimated for this. Later, we will restrict this parameter estimate to 1.

- Four additional variables (TRT, BASE, AGE and AVGBASE) are included in the Covariate box and treated as *inactive* as indicated by the symbol '< I >'. Thus, these variables will not affect the estimation of the parameters, but these variables will be cross-tabulated by class in the output.

Notice in the Classes box that the symbol '1-4' appears. This indicates that we will estimate a 1-class, 2-class, 3-class and a 4-class model.

- Click 'Estimate' to estimate these 4 models

After the estimation has completed:

- Click the data file name 'growth1.sav' to display the model summary table
- Right click in the model summary output table to retrieve the Model Summary Display
- Click to remove the checkmarks for L-square, df and p-value in the Model Summary Display

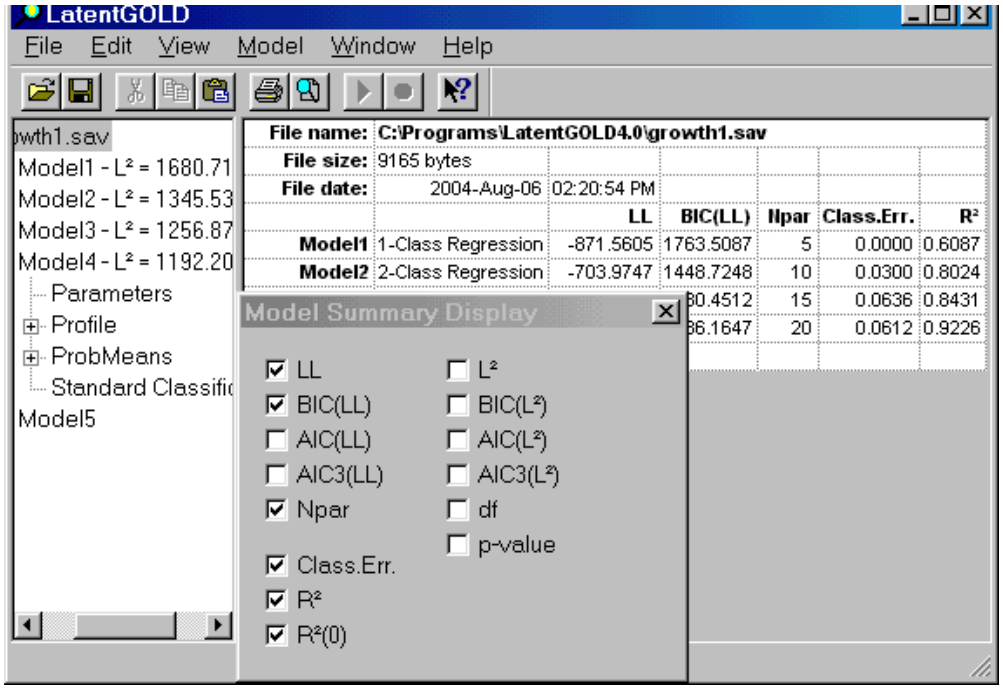


Figure 4. Model Summary Output and Model Summary Display

Notice that the BIC statistic is lowest for the 4-class model. We will examine the output for both the 3- and 4-class models here. The R^2 statistic is .84 for the 3-class and .92 for the 4-class model, and the misclassification error is approximately 6% for each of these models.

- Click 'Parameters' associated with the 4-class model

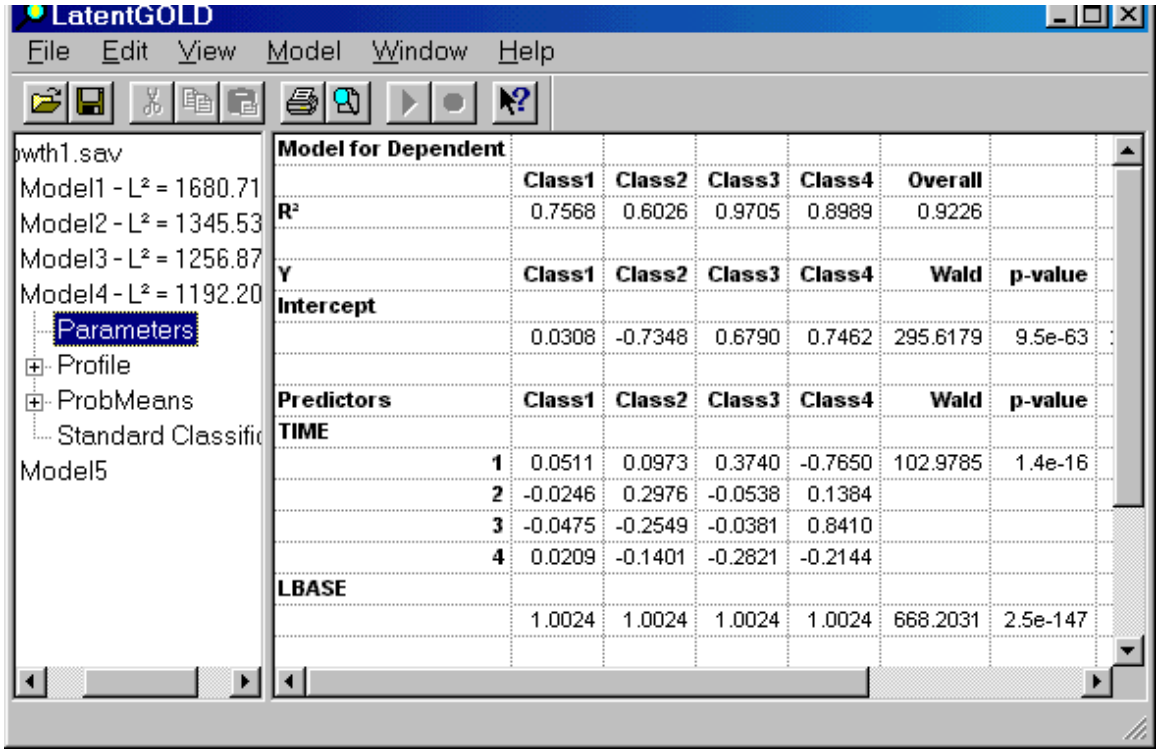


Figure 5. Parameters Output for 4-class Model

Note the following:

- the coefficient for LBASE is almost 1.
- the TIME estimates for class 1 are very close to zero, suggesting that this class shows no change from the baseline seizure rate. We will refer to the growth pattern for class 1 as 'no change'.
- The estimate for the Intercept (alpha) for class 2 is a large negative value suggesting an overall reduction in the seizure rate for this class. Despite the small setback between periods 1 and 2 (indicated by the beta increasing from .0973 to .2976) we will refer to this as the 'improved' class.
- Classes 3 and 4 show a large positive alpha, suggesting an overall increase (worsening) in the seizure rate.
- The betas for these classes show an abrupt increase in the number of seizures at some point during the follow-up period (time 1 for class 3 shows a beta of .3740; time 3 for class 4 shows a beta of .8410).
- We will see later that the cases classified into one of these 'unstable' classes are the same ones that were identified as 'outliers' in previous studies. We will refer to these classes as 'unstable'.

To display standard errors for these estimates, right click in the output table and select 'Standard Errors' from the popup menu.

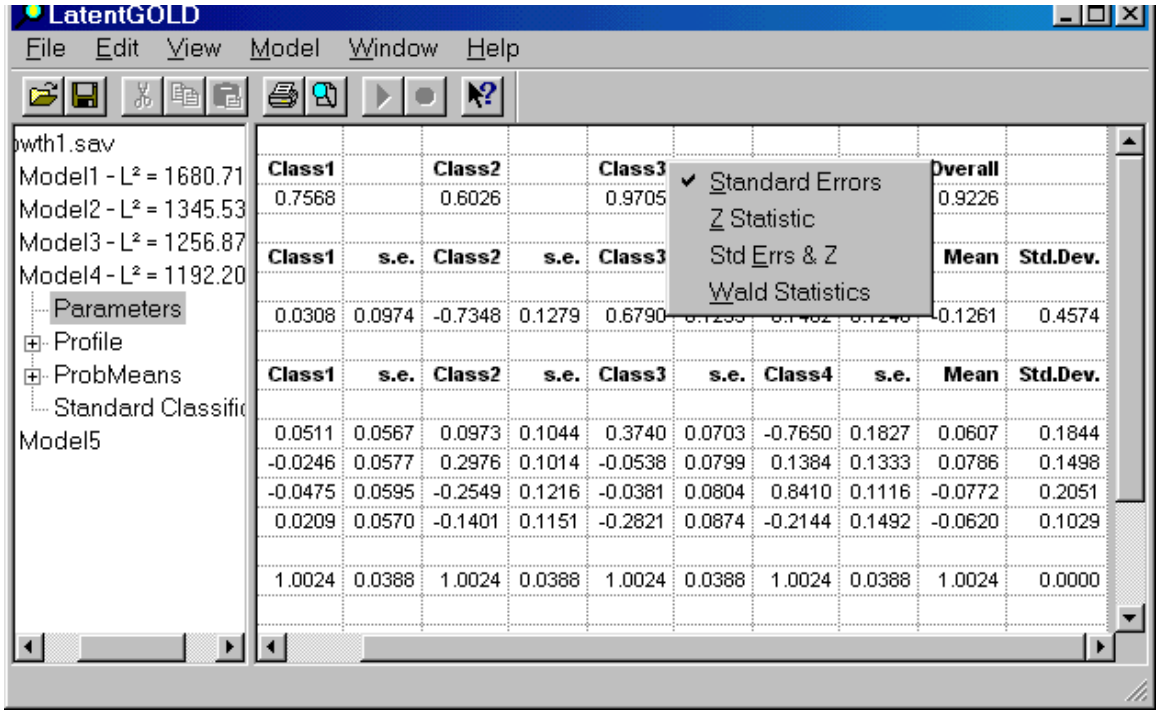


Figure 6. Parameters Output for 4-class model with Standard Errors

For class 1, the estimates for the alpha and beta parameters are all less than 1 standard deviation. Later, we will further refine class 1 by setting the TIME estimates (betas) to zero. The estimates for alpha for the other classes are well above 2 standard errors.

Note that the two right-most columns provide the *Overall* mean and standard deviation for the alpha and beta parameters. For LBASE the standard error for the *Overall* effect is 0 because this is a fixed effect (i.e., it is class independent). The other estimates have non-zero standard errors because they differ depending upon the class (i.e., they are treated as *class dependent* or *random effects*).

- Click on 'Profile' to display the Profile output for the 4-class model.

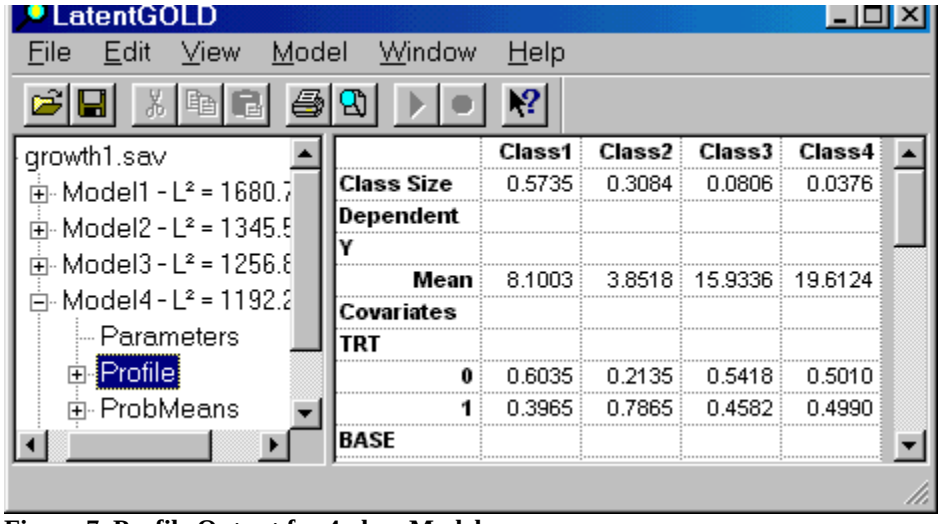
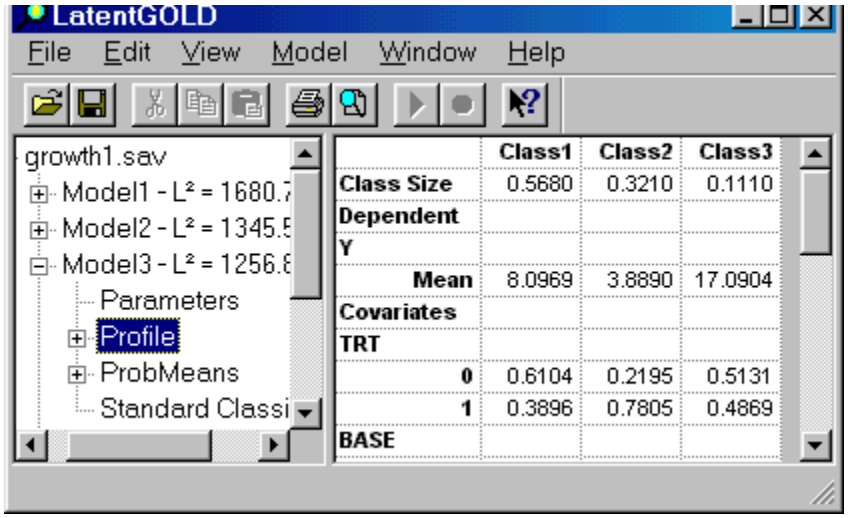


Figure 7. Profile Output for 4-class Model

The row labeled 'Class Size' indicates that Class 1 (the 'no change' class) represents 57% of the cases. Class 2 (those who improved) contains about 31% of the cases. The remaining 12% of the cases (classes 3 and 4) are the unstable ones.

- Click on the icon to the left of Model 3 to expand it
- Click on Profile to open the Profile output for the 3-class model



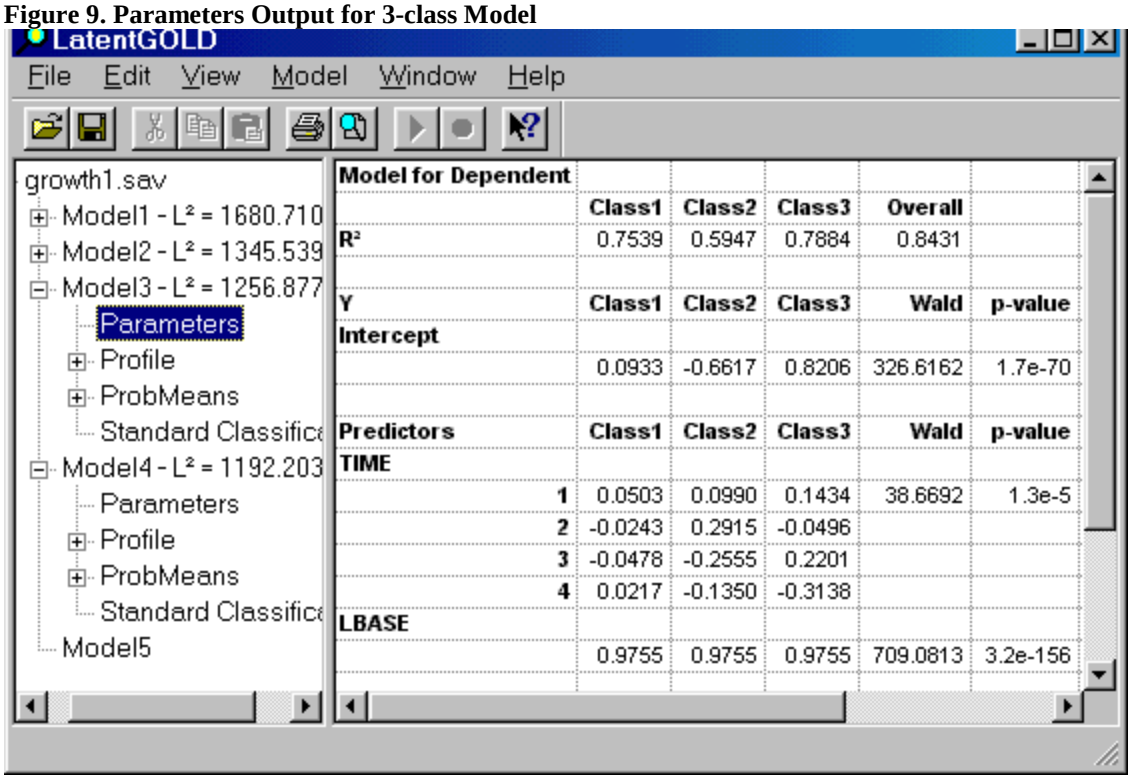
	Class1	Class2	Class3
Class Size	0.5680	0.3210	0.1110
Dependent			
Y			
Mean	8.0969	3.8890	17.0904
Covariates			
TRT			
0	0.6104	0.2195	0.5131
1	0.3896	0.7805	0.4869
BASE			

Figure 8. Profile Output for 3-class Model.

Notice that the class sizes for classes 1 and 2 are about the same as the corresponding classes for the 4-class model.

- Click Parameters to display the Parameters output for the 3-class model

Figure 9. Parameters Output for 3-class Model



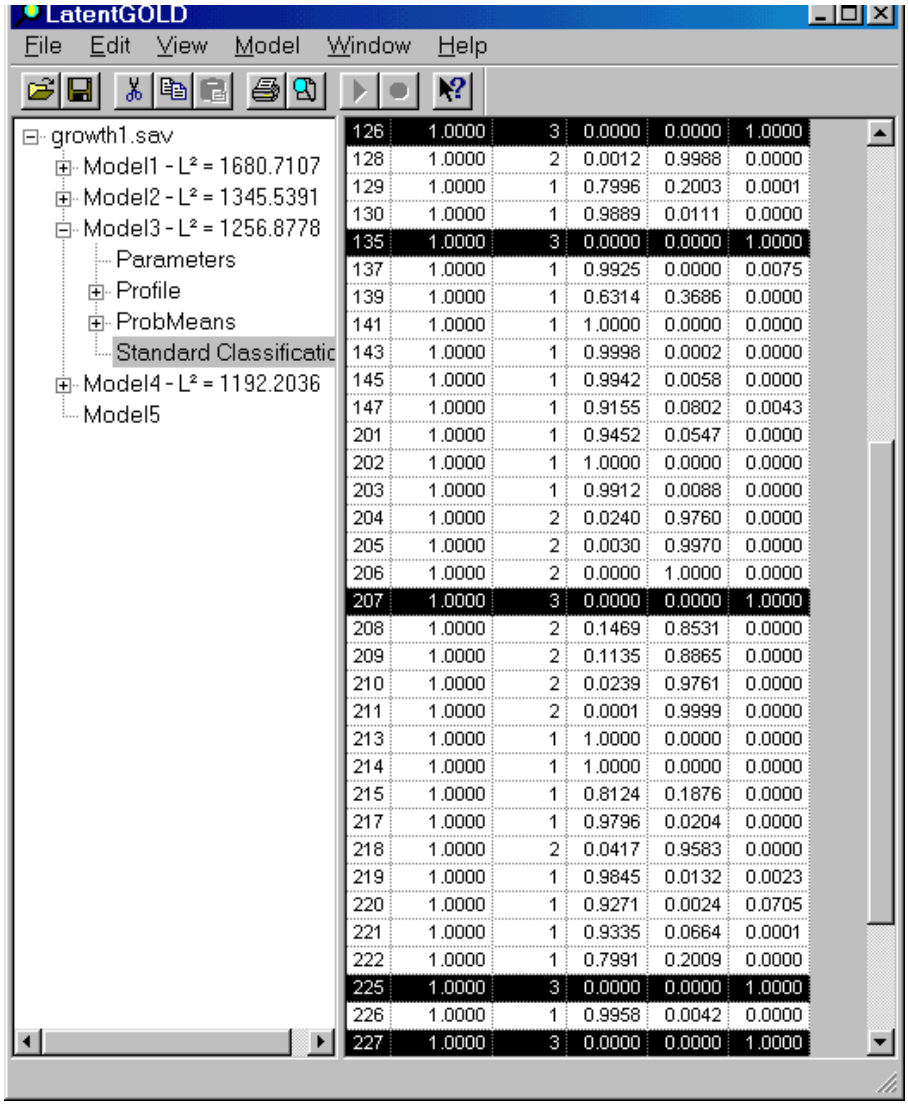
Model for Dependent					
	Class1	Class2	Class3	Overall	
R^2	0.7539	0.5947	0.7884	0.8431	
Y	Class1	Class2	Class3	Wald	p-value
Intercept					
	0.0933	-0.6617	0.8206	326.6162	1.7e-70
Predictors	Class1	Class2	Class3	Wald	p-value
TIME					
1	0.0503	0.0990	0.1434	38.6692	1.3e-5
2	-0.0243	0.2915	-0.0496		
3	-0.0478	-0.2555	0.2201		
4	0.0217	-0.1350	-0.3138		
LBASE					
	0.9755	0.9755	0.9755	709.0813	3.2e-156

Notice that the alpha and beta estimates for classes 1 and 2 are also similar to those for the 4-class solution. Thus, we see that the classes 1 and 2 are basically identical in the 3- and 4-class solutions.

To see that Class 3 contains both unstable classes from the 4-class solution:

- Click on Standard Classification Output for the 3-class model.
- Scroll down to identify the cases for which Modal = 3

These cases are #112, #126, #135, #207, #225, and #227. (The posterior membership probabilities for five of these cases are shown below):



126	1.0000	3	0.0000	0.0000	1.0000
128	1.0000	2	0.0012	0.9988	0.0000
129	1.0000	1	0.7996	0.2003	0.0001
130	1.0000	1	0.9889	0.0111	0.0000
135	1.0000	3	0.0000	0.0000	1.0000
137	1.0000	1	0.9925	0.0000	0.0075
139	1.0000	1	0.6314	0.3686	0.0000
141	1.0000	1	1.0000	0.0000	0.0000
143	1.0000	1	0.9998	0.0002	0.0000
145	1.0000	1	0.9942	0.0058	0.0000
147	1.0000	1	0.9155	0.0802	0.0043
201	1.0000	1	0.9452	0.0547	0.0000
202	1.0000	1	1.0000	0.0000	0.0000
203	1.0000	1	0.9912	0.0088	0.0000
204	1.0000	2	0.0240	0.9760	0.0000
205	1.0000	2	0.0030	0.9970	0.0000
206	1.0000	2	0.0000	1.0000	0.0000
207	1.0000	3	0.0000	0.0000	1.0000
208	1.0000	2	0.1469	0.8531	0.0000
209	1.0000	2	0.1135	0.8865	0.0000
210	1.0000	2	0.0239	0.9761	0.0000
211	1.0000	2	0.0001	0.9999	0.0000
213	1.0000	1	1.0000	0.0000	0.0000
214	1.0000	1	1.0000	0.0000	0.0000
215	1.0000	1	0.8124	0.1876	0.0000
217	1.0000	1	0.9796	0.0204	0.0000
218	1.0000	2	0.0417	0.9583	0.0000
219	1.0000	1	0.9845	0.0132	0.0023
220	1.0000	1	0.9271	0.0024	0.0705
221	1.0000	1	0.9335	0.0664	0.0001
222	1.0000	1	0.7991	0.2009	0.0000
225	1.0000	3	0.0000	0.0000	1.0000
226	1.0000	1	0.9958	0.0042	0.0000
227	1.0000	3	0.0000	0.0000	1.0000

Figure 10. Standard Classification Output for the 3-class Model

- Compare this with the Standard Classification Output for the 4-class model

Notice that the 6 cases assigned to class 3 in the 3-class solution have posterior membership probabilities of 1.000 of being in class 3. These cases are the same as those assigned to classes 3 or class 4 in the 4 class solution, again with posterior probabilities equal to 1.000.

These 6 cases are the same as those identified as outliers in the analyses of these data by Rabe-Hesketh and Skrondal (see http://fmwww.bc.edu/repec/usug2003/diag.pdf) and excluded from their study.

While strict adherence to the BIC criteria would cause us to conclude that there are more than 4 distinct time trends for these data, for our purposes in this tutorial, we will focus here on the 3 class solution which identifies 3 classes that are of substantive interest. Class 1 shows no change, class 2 shows substantial improvement and class 3 contains a small number of unstable outliers who show a substantial increase at some follow-up time period. From a substantive perspective, we will be interested in determining to what extent treatment with the drug Progabide vs. the Placebo can explain the growth pattern exhibited by class 2 as opposed to that exhibited by class 1.

We will now refine the 3-class solution by applying the following restrictions:

- Restrict the beta for LBASE to 1
- Restrict the betas for TIME to 0 for class 1 (note that we might also choose to set alpha to 0 for this class)

- Double-click on Model 3
- Click on Model Tab

From the Model Tab,

- Select the row of 1s associated with LBASE
- Right-click to retrieve the restrictions popup menu
- Select 'Offset'

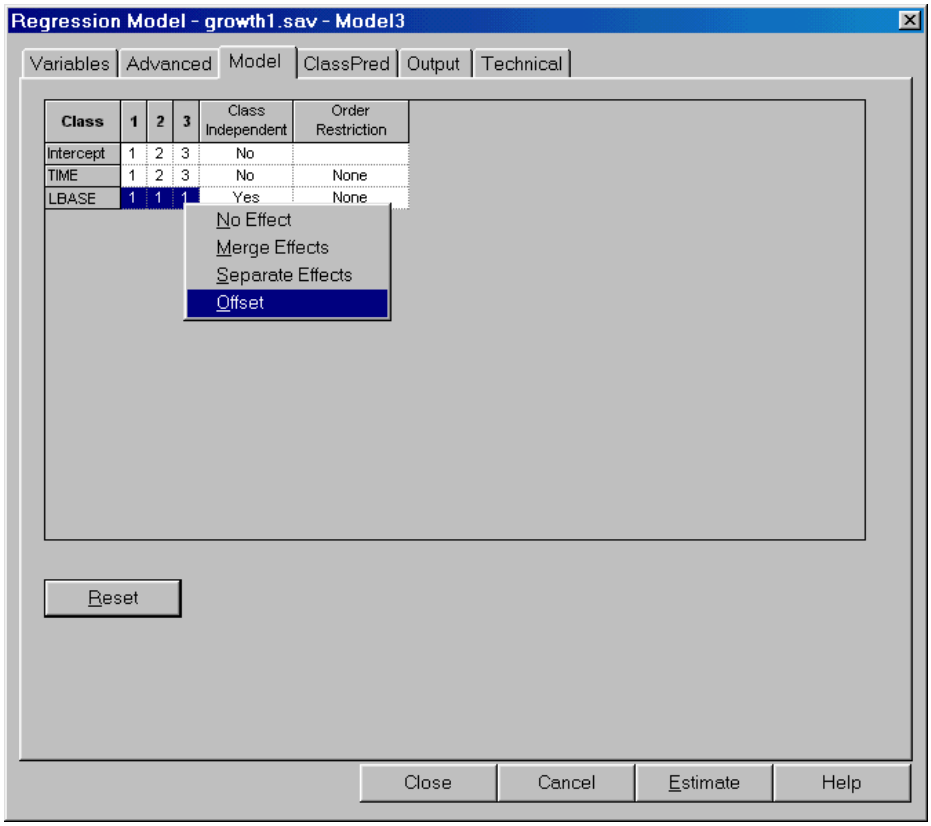


Figure 11. Model Tab for 3-class Model

The '1's change to '*'s.

To implement the zero restriction:

- Select the '1' associated with the Class 1 TIME effect
- Right-click to retrieve the restrictions popup menu
- Select 'No Effect'

The one changes to the symbol '-' to indicate that this effect is restricted to 0.

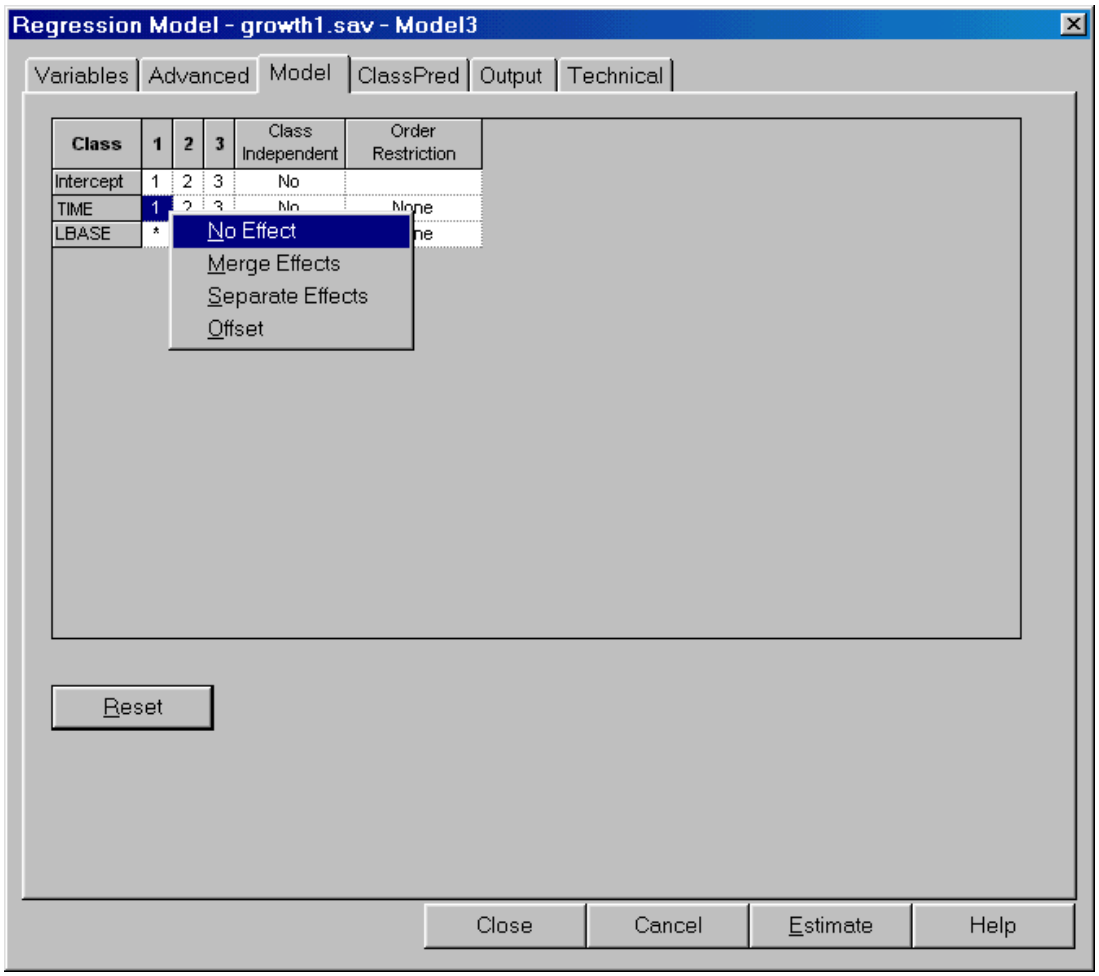
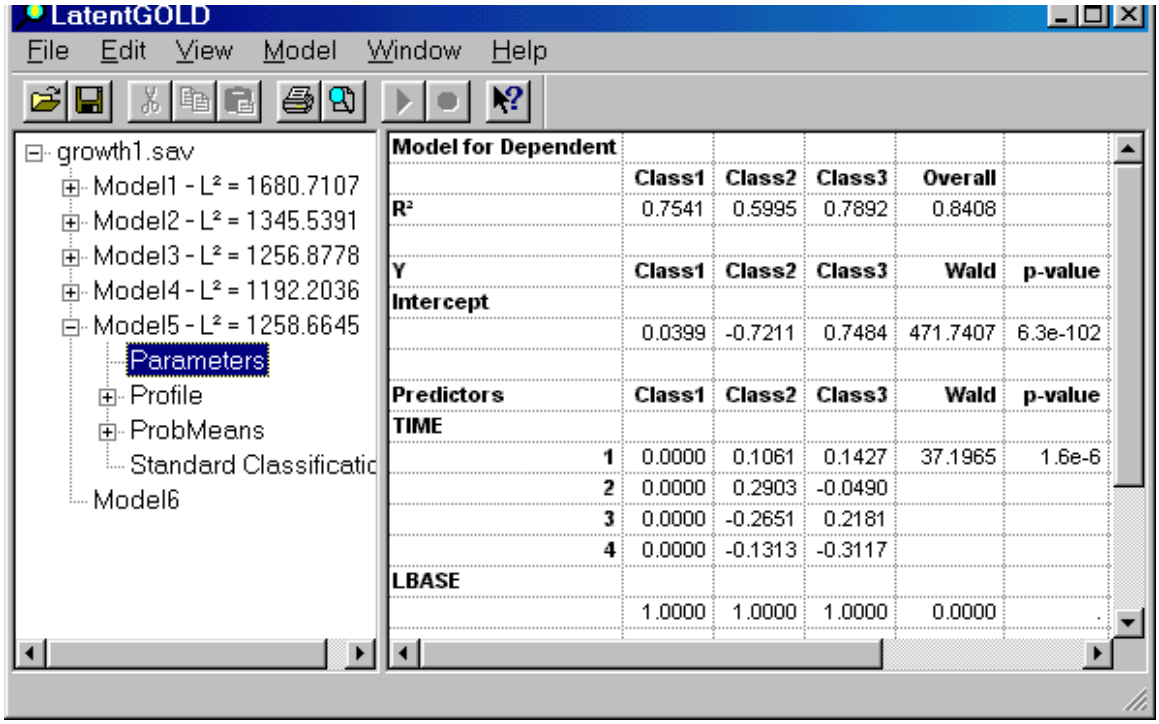


Figure 12. Implementing the zero restriction

- Click Estimate
- Click Parameters to display the parameters output



Model for Dependent					
	Class1	Class2	Class3	Overall	
R^2	0.7541	0.5995	0.7892	0.8408	
Y	Class1	Class2	Class3	Wald	p-value
Intercept					
	0.0399	-0.7211	0.7484	471.7407	6.3e-102
Predictors	Class1	Class2	Class3	Wald	p-value
TIME					
1	0.0000	0.1061	0.1427	37.1965	1.6e-6
2	0.0000	0.2903	-0.0490		
3	0.0000	-0.2651	0.2181		
4	0.0000	-0.1313	-0.3117		
LBASE					
	1.0000	1.0000	1.0000	0.0000	.

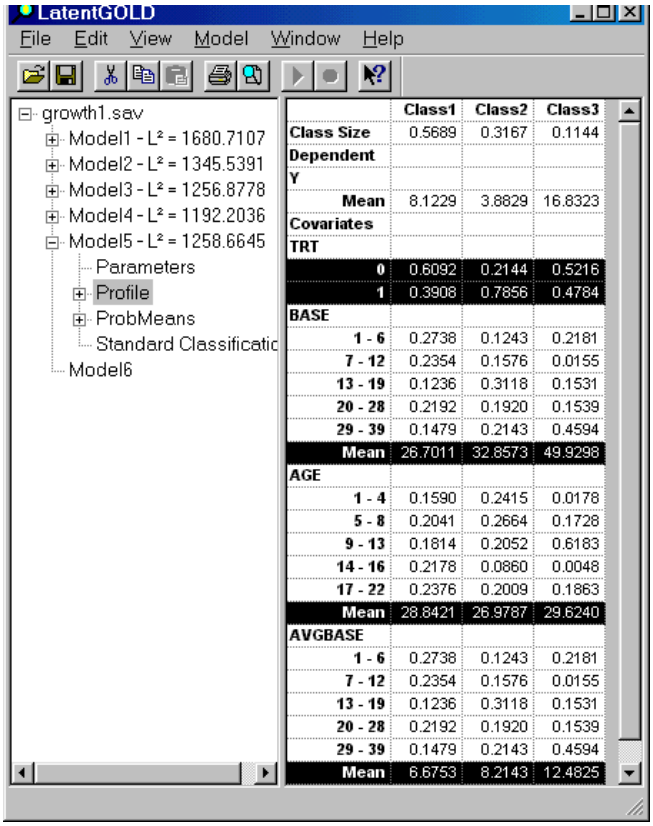
Figure 13. Parameters Output for new 3-class Model

The estimates are similar to Model 3 except that the restrictions have been incorporated into the model.

To examine the treatment effect:

- Click Profile

Figure 14. Profile Output

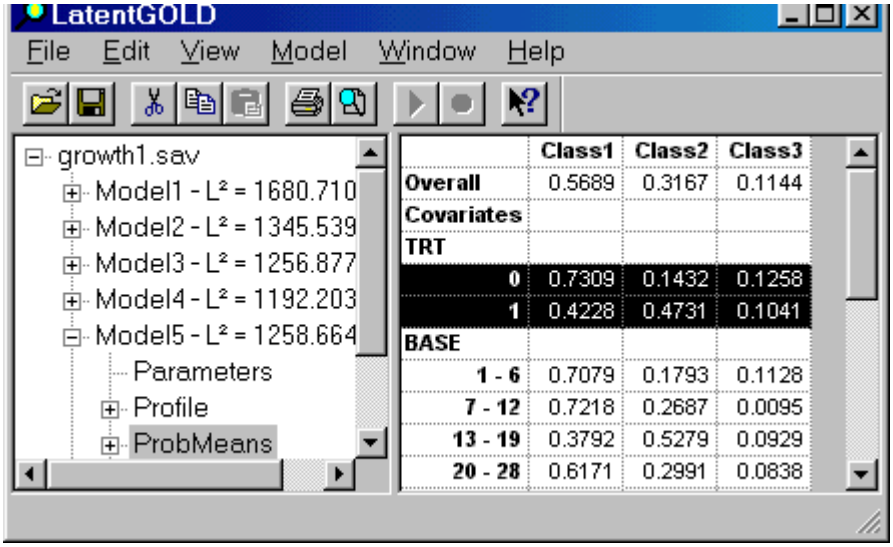


	Class1	Class2	Class3
Class Size	0.5689	0.3167	0.1144
Dependent			
Y			
Mean	8.1229	3.8829	16.8323
Covariates			
TRT			
0	0.6092	0.2144	0.5216
1	0.3908	0.7856	0.4784
BASE			
1 - 6	0.2738	0.1243	0.2181
7 - 12	0.2354	0.1576	0.0155
13 - 19	0.1236	0.3118	0.1531
20 - 28	0.2192	0.1920	0.1539
29 - 39	0.1479	0.2143	0.4594
Mean	26.7011	32.8573	49.9298
AGE			
1 - 4	0.1590	0.2415	0.0178
5 - 8	0.2041	0.2664	0.1728
9 - 13	0.1814	0.2052	0.6183
14 - 16	0.2178	0.0860	0.0048
17 - 22	0.2376	0.2009	0.1863
Mean	28.8421	26.9787	29.6240
AVGBASE			
1 - 6	0.2738	0.1243	0.2181
7 - 12	0.2354	0.1576	0.0155
13 - 19	0.1236	0.3118	0.1531
20 - 28	0.2192	0.1920	0.1539
29 - 39	0.1479	0.2143	0.4594
Mean	6.6753	8.2143	12.4825

These column proportions show that most of those in the 'No Change' class received the placebo while most of those in the 'Improved' class received the drug treatment. From the means associated with AVGBASE we also see that the 'Unstable' group had a substantially larger base rate (12.4825) than the other classes.

To display the corresponding row proportions:

- Click ProbMeans



	Class1	Class2	Class3
Overall	0.5689	0.3167	0.1144
Covariates			
TRT			
0	0.7309	0.1432	0.1258
1	0.4228	0.4731	0.1041
BASE			
1 - 6	0.7079	0.1793	0.1128
7 - 12	0.7218	0.2687	0.0095
13 - 19	0.3792	0.5279	0.0929
20 - 28	0.6171	0.2991	0.0838

Figure 15. ProbMeans Output

We see that of those who received the drug, 47.31% are in the 'Improved' class compared to only 14.32% of those who received the placebo. Similarly, only 42.28% showed 'No change' compared to 73.09% of the placebo group. These numbers are in the direction expected if the drug reduced the seizure rate.

To test a slightly different treatment effect by using Latent GOLD with an active covariate, see Tutorial #7B.

To rename this model for future use:

- Click 'Model 5' to select it

Model 5 is now highlighted. To enter the Edit Model

- Click the highlighted 'Model 5'
- Replace 'Model 5' by typing "Final 3-class model"

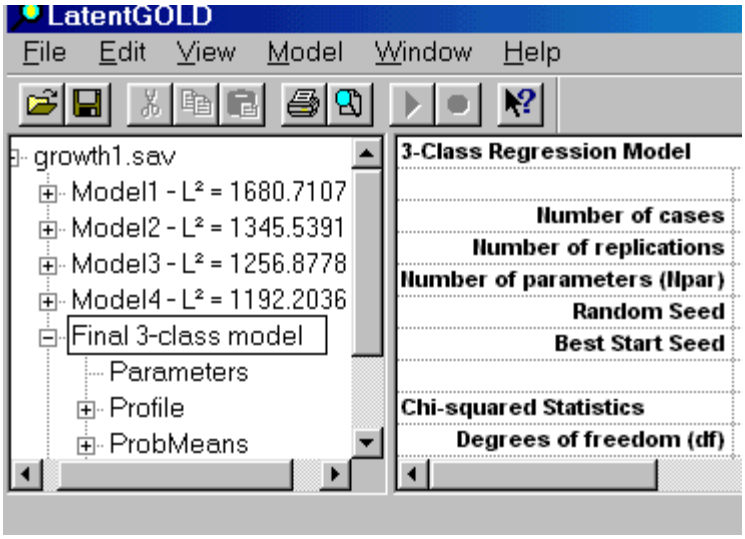


Figure 16. Renaming Model 5

To save this model for future use:

- Click 'Final 3-class model' to select it

From the File Menu

- Select 'Save Definition'
- Click Save

The model definition is saved as the file 'Final 3-class model.lgf'