

# Using Mixture Latent Markov Models for Analyzing Change with Longitudinal Data

Jay Magidson, Ph.D.  
President, Statistical Innovations Inc.  
Belmont, MA., U.S.

Presented at Modern Modeling Methods (M3) 2013, University of Connecticut

# Abstract

Mixture latent Markov (MLM) models are latent class models containing both time-constant and time-varying discrete latent variables. We introduce and illustrate the utility of latent Markov models in 3 real world examples using the new GUI in Latent GOLD<sup>®</sup> 5.0.

MLM models often fit longitudinal data better than comparable mixture latent growth models because of the inclusion of transition probabilities as model parameters which directly account for first-order autocorrelation in the data. Lack of fit can be localized (e.g., first-order, second-order autocorrelation) using new longitudinal bivariate residuals (L-BVRs) to suggest model modification.

# Outline of Presentation

- Introduction to latent Markov (LM) and mixture latent Markov (MLM) Models\*
- Example 1: LM model with time-homogeneous transition probabilities
- Example 2: MLM model with time-heterogeneous transitions
- Example 3: MLM model with time-heterogeneous transitions and covariates

In each of these examples latent Markov outperformed the corresponding latent growth model, the largest improvement being in the Lag1 BVRs (i.e., first-order autocorrelation).

\* (Mixture) *latent* Markov models are also known as (mixture) *hidden* Markov models

# Latent (Hidden) Markov Models are Defined by 3 Sets of Equations

- **Initial State probs:**  $P(X_0 = s)$  – categories of  $X$  are called *latent states* (rather than *latent classes*) since a person may change from one state to another over time.

**Logit model** may include covariates  $Z$  (e.g., AGE, SEX)  $\log \frac{P(x_0 = s)}{P(x_0 = 1)} = \alpha_{0s}$

- **Transition probs:**  $P(X_t = r | X_{t-1} = s)$   $\log \frac{P(x_t = r | x_{t-1} = s)}{P(x_t = 1 | x_{t-1} = s)} = \gamma_{0r} + \gamma_{1rs} + \gamma_{2rt} + \gamma_{3rst}$

**Logit** may include time-varying and fixed predictors

- **Measurement model probs:**  $P(y_t = j | X_t = s)$

One or more dependent variables  $Y$ , of possibly different scale types (e.g., continuous, count, dichotomous, ordinal, nominal)

$$\log \frac{P(y_{it} = \ell | x_t = s)}{P(y_{it} = 1 | x_t = s)} = \beta_{0\ell} + \beta_{1\ell s}$$

\* Equations can be customized using Latent GOLD 5.0 GUI and/or syntax.

# Equations:

Latent Markov (LM): 
$$P(\mathbf{y}_i | \mathbf{z}_i) = \sum_{x_0=1}^K \sum_{x_1=1}^K \dots \sum_{x_T=1}^K P(x_0, x_1, \dots, x_T | \mathbf{z}_i) P(\mathbf{y}_i | x_0, x_1, \dots, x_T, \mathbf{z}_i)$$

Initial latent state &  
Transition sub-models 
$$P(x_0, x_1, \dots, x_T | \mathbf{z}_i) = P(x_0 | \mathbf{z}_{i0}) \prod_{t=1}^T P(x_t | x_{t-1}, \mathbf{z}_{it})$$

Measurement  
sub-model 
$$P(\mathbf{y}_i | x_0, x_1, \dots, x_T, \mathbf{z}_i) = \prod_{t=0}^T P(\mathbf{y}_{it} | x_t, \mathbf{z}_{it}) = \prod_{t=0}^T \prod_{j=1}^J P(y_{itj} | x_t, \mathbf{z}_{it})$$

## Mixture Latent Markov (MLM):

$$P(\mathbf{y}_i | \mathbf{z}_i) = \sum_{w=1}^L \sum_{x_0=1}^K \sum_{x_1=1}^K \dots \sum_{x_T=1}^K P(w, x_0, x_1, \dots, x_T | \mathbf{z}_i) P(\mathbf{y}_i | w, x_0, x_1, \dots, x_T, \mathbf{z}_i)$$

$$P(w, x_0, x_1, \dots, x_T | \mathbf{z}_i) = P(w | \mathbf{z}_i) P(x_0 | w, \mathbf{z}_{i0}) \prod_{t=1}^T P(x_t | x_{t-1}, w, \mathbf{z}_{it})$$

$$P(\mathbf{y}_i | w, x_0, x_1, \dots, x_T, \mathbf{z}_i) = \prod_{t=0}^T P(\mathbf{y}_{it} | x_t, w, \mathbf{z}_{it}) = \prod_{t=0}^T \prod_{j=1}^J P(y_{itj} | x_t, w, \mathbf{z}_{it})$$

# Relationship between Mixture Latent Markov and Mixture Latent Growth Models

Classification of latent class models for longitudinal research

	Model name	Transition structure	Unobserved heterogeneity	Measurement error
I	Mixture latent Markov	yes	yes	yes
II	Mixture Markov	yes	yes	no
III	Latent Markov	yes	no	yes
IV	Standard Markov*	yes	no	no
V	Mixture latent growth	no	yes	yes
VI	Mixture growth	no	yes	no
VII	Standard latent class	no	no	yes
VIII	Independence*	no	no	no

\*This model is not a latent class model.

Vermunt, Tran and Magidson (2008) “Latent class models in longitudinal research”, chapter 23 in Handbook of Longitudinal Research, S. Menard Editor, Academic Press.

# Longitudinal Bivariate Residuals (L-BVRs)

**Longitudinal bivariate residuals** quantify for each response variable  $Y_k$  how well the overall trend as well as the first- and second-order autocorrelations are predicted by the model.

$$\text{BVR.Time} = \text{BVR}_k(\text{time}, y_k),$$

$$\text{BVR.Lag1} = \text{BVR}_k(y_k[t-1], y_k[t])$$

residual autocorrelations remaining unexplained by the model.

$$\text{and BVR.Lag2} = \text{BVR}_k(y_k[t-2], y_k[t])$$

$$\text{Lag1: } \text{BVR}_k = \sum_r \sum_s \frac{(n_{krs} - m_{ksr})^2}{m_{krs}}$$

where  $m_{ksr}$  is based on  $P(Y_{kit-1} = r, Y_{kit} = s)$ , that is

$$m_{ksr} = \sum_i \sum_t P(Y_{kit-1} = r, Y_{kit} = s).$$

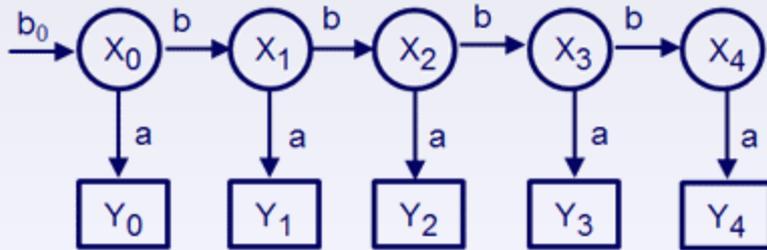
$$= P(Y_{kit-1} = r | X_{it-1} = x_{t-1}) P(Y_{kit} = s | X_{it} = x_t)$$

# Example 1: Loyalty Data in Long File Format

	id	y	freq	time
1	1	1	352	1
2	1	1	352	2
3	1	1	352	3
4	1	1	352	4
5	1	1	352	5
6	2	0	44	1
7	2	1	44	2
8	2	1	44	3
9	2	1	44	4
10	2	1	44	5

- N=631 respondents
- T= 5 time points
- Dichotomous Y:
  - Choose Brand A?
    - 1=Yes
    - 0=No
- $Y=(Y_1, Y_2, Y_3, Y_4, Y_5)$
- $2^5 = 32$  response patterns id=1,2,...,32
- 'freq' is used as a case weight

# Loyalty Model with Time-homogeneous Transitions



Measurement equivalence

LatentGOLD

File Edit View Model Window Help

latent Markov -  $L^2 = 20.4902$

- Parameters
  - Profile
  - ProbMeans-Posterior
  - Freqs/Residuals
  - Bivariate Residuals
  - Classification-Posterior
  - EstimatedValues-Regression
  - EstimatedValues-Model**
- Model4

poulsen.wideformat.sav

		State[=0]	
		1	2
		0.7515	0.2485
		State	
State[-1]	State	1	2
1	1	0.9763	0.0237
2	1	0.1956	0.8044
		y	
State		other brand	brand A
1		0.0476	0.9524
2		0.7720	0.2280

Initial State Probability ( $b_0$ )

Transition Probabilities ( $b$ )

Measurement Model Probabilities ( $a$ )

# Example 1: LM Model Outperforms Latent Growth Model

2-state time-homogeneous latent Markov model  
fits well: ( $p=.77$ ), small L-BVRs

2-class latent growth model (“2-class Regression”) is rejected ( $p=.0084$ )

The screenshot shows the LatentGOLD software window with a menu bar (File, Edit, View, Model, Window, Help) and a toolbar. The main window displays a list of models on the left and a table of fit statistics on the right.

		Npar	L <sup>2</sup>	df	p-value
null	1-Class Regression	5	504.6663	26	3.8e-90
2-class Regression	2-Class Regression	11	45.8639	20	0.00084
latent Markov	2-State 1-Class	5	20.4902	26	0.77

L-BVR	2-class Regression	LM
Time	0.000	.0785
Lag1	8.739* (p = .003)	.0115
Lag2	0.065	.3992

Lag1 BVR pinpoints problem in LC growth model as failure to explain 1<sup>st</sup> order autocorrelation

# Example 2: Life Satisfaction Model with Time-Heterogeneous Transitions

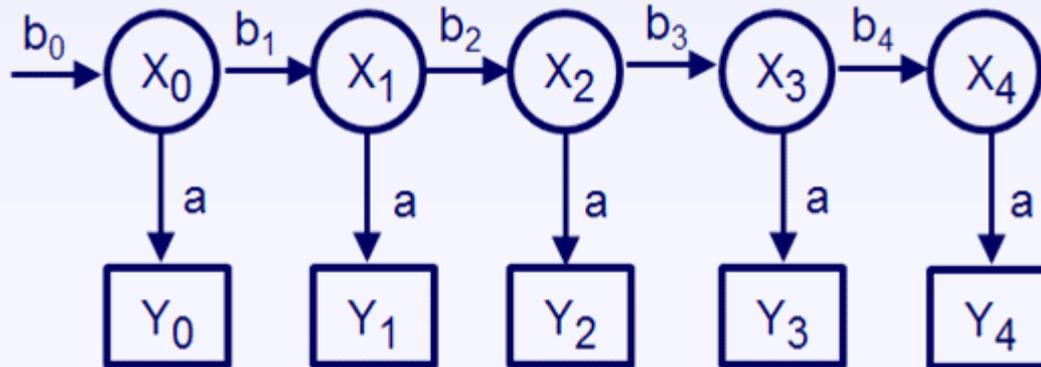
	id	time	response	weight
1	1	1	1	891
2	1	2	1	891
3	1	3	1	891
4	1	4	1	891
5	1	5	1	891
6	2	1	1	176
7	2	2	1	176
8	2	3	1	176
9	2	4	1	176
10	2	5	2	176

- N=5,147 respondents
- T= 5 time points
- Dichotomous Y:
  - Satisfied with life?
    - 1=No
    - 2=Yes
- $Y=(Y_1, Y_2, Y_3, Y_4, Y_5)$
- $2^5 = 32$  response patterns id=1,2,...,32
- 'weight' is used as a case weight

## Models:

- Null
- Time heterogeneous LM
- 2-class mixture LM
- Restricted (Mover-Stayer)

## Ex. 2: Time-heterogeneous Models Latent Markov Model



Separate transition probability parameters exist for each pair of adjacent time points

# Example 2: Model Parameters Time-heterogeneous Model

LatentGOLD

File Edit View Model Window Help

satisfaction.sav

- null -  $L^2 = 5026.1237$
- 2-state time heterogeneous
  - Parameters
  - Profile
  - Profile-Longitudinal
  - ProbMeans-Posterior
  - Bivariate Residuals
  - EstimatedValues-Regre
  - EstimatedValues-Model3

		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

Initial State Probability ( $b_0$ )

Transition Probabilities ( $b_t$ )

Measurement Model Probabilities ( $a$ )

Estimated values for 2-state LM model

## Ex. 2: Mover-Stayer Time-heterogeneous Latent Markov Model

Both the unrestricted and Mover-Stayer 2-class MLM models fit well ( $p=.95$  and  $.71$ ), the BIC statistic preferring the Mover-Stayer model.

The screenshot shows the LatentGOLD software window with a menu bar (File, Edit, View, Model, Window, Help) and a toolbar. The main area displays a table of model fit statistics for the file 'satisfaction.sav'. The table includes columns for model name, Npar, L<sup>2</sup>, df, and p-value. The models listed are: null (L<sup>2</sup> = 5026.1237), 2-class Regression (L<sup>2</sup> = 275.34), 2-class LM (L<sup>2</sup> = 14.3077), and 2-class LM Mover-Stayer (L<sup>2</sup> = ...). The p-values for the 2-class LM and 2-class LM Mover-Stayer models are 0.71 and 0.95, respectively.

		Npar	L <sup>2</sup>	df	p-value
null	1-Class Regression	5	5026.1237	26	5.2e-1060
2-class Regression	2-Class Regression	11	275.3405	20	8.5e-47
2-class LM	2-State 2-Class	13	14.3077	18	0.71
2-class LM Mover-Stayer	2-State 2-Class	15	7.9180	16	0.95

L-BVR	2-class Regression	2-class LM	2-class LM Mover-Stayer
Time	0.0	0.02	0.03
Lag1	55.1 (p=1.1E-13)	0.00	0.00
Lag2	1.7	0.10	1.97

# Mover-Stayer Time-heterogeneous 2-class MLM Model

Estimated Values output for 2-state time-heterogeneous MLM model with 2 classes

LatentGOLD

File Edit View Model Window Help

satisfaction.sav

- [-] null -  $L^2 = 5026.1237$
- [-] 2-state time heterogeneous
- [-] 2-class LM -  $L^2 = 7.9180$ 
  - [-] Parameters
  - [-] Profile
  - [-] Profile-Longitudinal
  - [-] ProbMeans-Posterior
  - [-] Bivariate Residuals
  - [-] EstimatedValues-Regres
  - [-] **EstimatedValues-Model**
  - [-] Iteration Detail
- [-] Model4

		Class	
		1	2
		0.6313	0.3687

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

		State	
		1	2
time	Class	State[-1]	
1987	1	1	0.9591
1987	1	2	0.0197
1987	2	1	0.4175
1987	2	2	0.3879
1988	1	1	0.9580
1988	1	2	0.0193
1988	2	1	0.4108
1988	2	2	0.3834
1989	1	1	0.9707
1989	1	2	0.0223
1989	2	1	0.5027
1989	2	2	0.4178
1990	1	1	0.9839
1990	1	2	0.0277
1990	2	1	0.6514
1990	2	2	0.4728

		response	
		dissatisfied	satisfied
1		0.1085	0.8915
2		0.9343	0.0657

Class Size

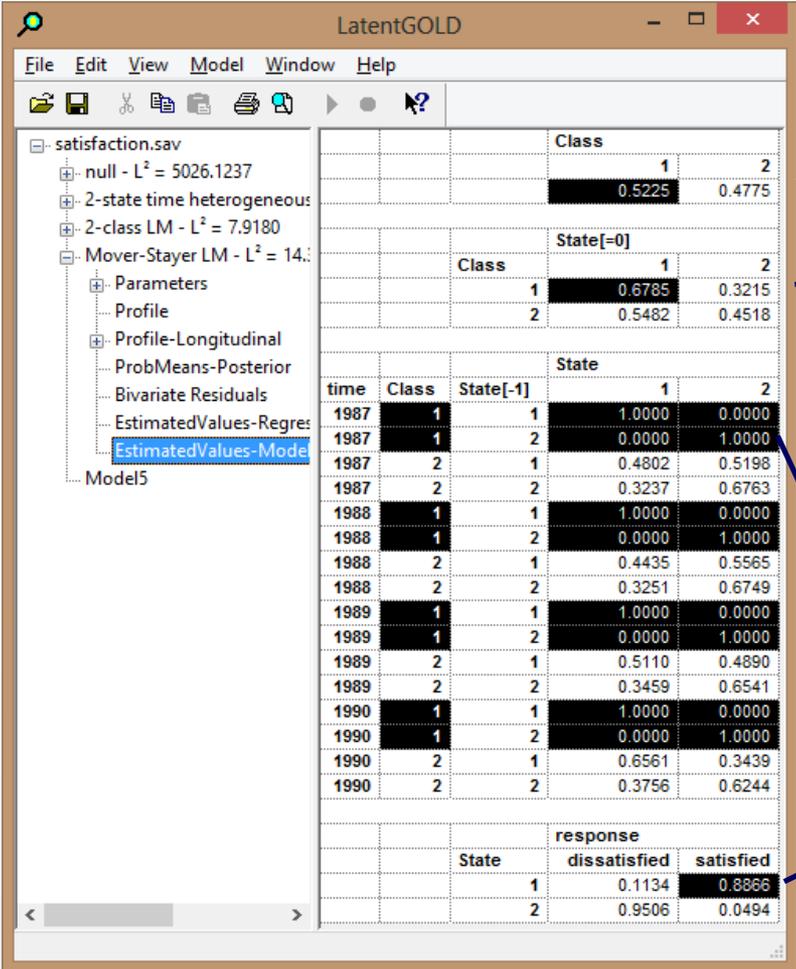
Initial State

Transition Probabilities

Measurement model

# Example 2: Mover-Stayer Time-heterogeneous Latent Markov Model

The Estimated Values output 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.



		Class	
		1	2
		0.5225	0.4775

		State[=0]	
		1	2
1	0.6785	0.3215	
2	0.5482	0.4518	

time	Class	State[=1]	State	1	2
1987	1	1	1.0000	0.0000	
1987	1	2	0.0000	1.0000	
1987	2	1	0.4802	0.5198	
1987	2	2	0.3237	0.6763	
1988	1	1	1.0000	0.0000	
1988	1	2	0.0000	1.0000	
1988	2	1	0.4435	0.5565	
1988	2	2	0.3251	0.6749	
1989	1	1	1.0000	0.0000	
1989	1	2	0.0000	1.0000	
1989	2	1	0.5110	0.4890	
1989	2	2	0.3459	0.6541	
1990	1	1	1.0000	0.0000	
1990	1	2	0.0000	1.0000	
1990	2	1	0.6561	0.3439	
1990	2	2	0.3756	0.6244	

		response	
		dissatisfied	satisfied
1	0.1134	0.8866	
2	0.9506	0.0494	

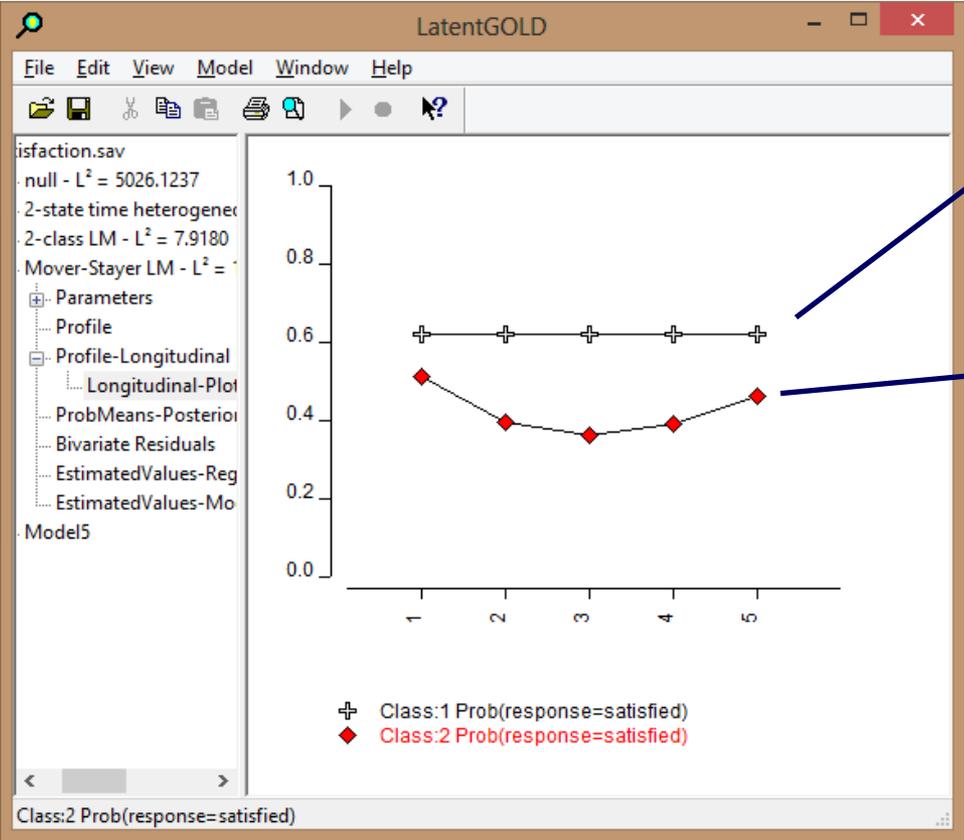
Class Size

Initial State

Transition Probabilities  
– note that class 1 probability of staying in the same state has been restricted to 1.

Measurement model

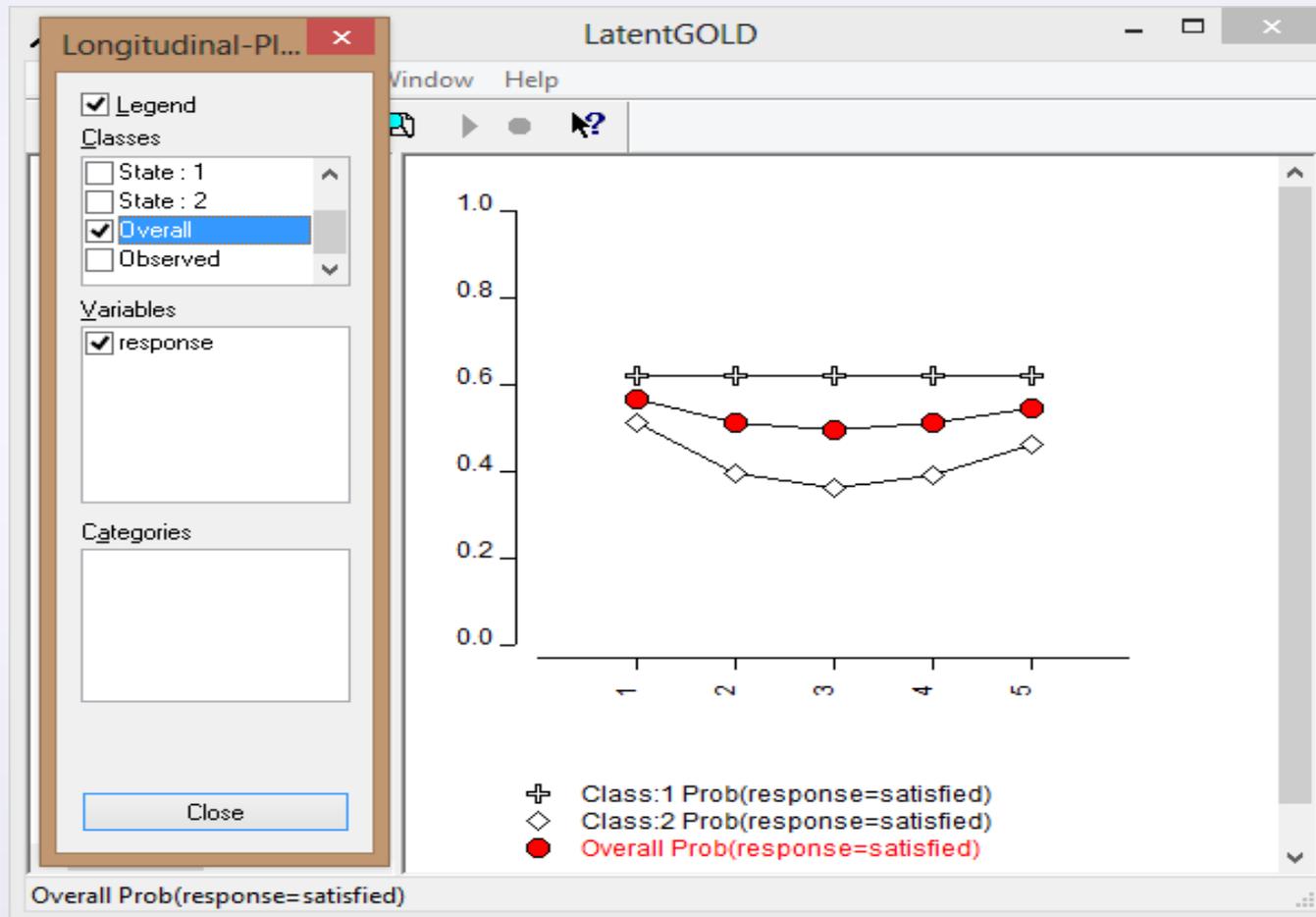
# Example 2: Longitudinal Profile Plot for the Mover-Stayer LM model



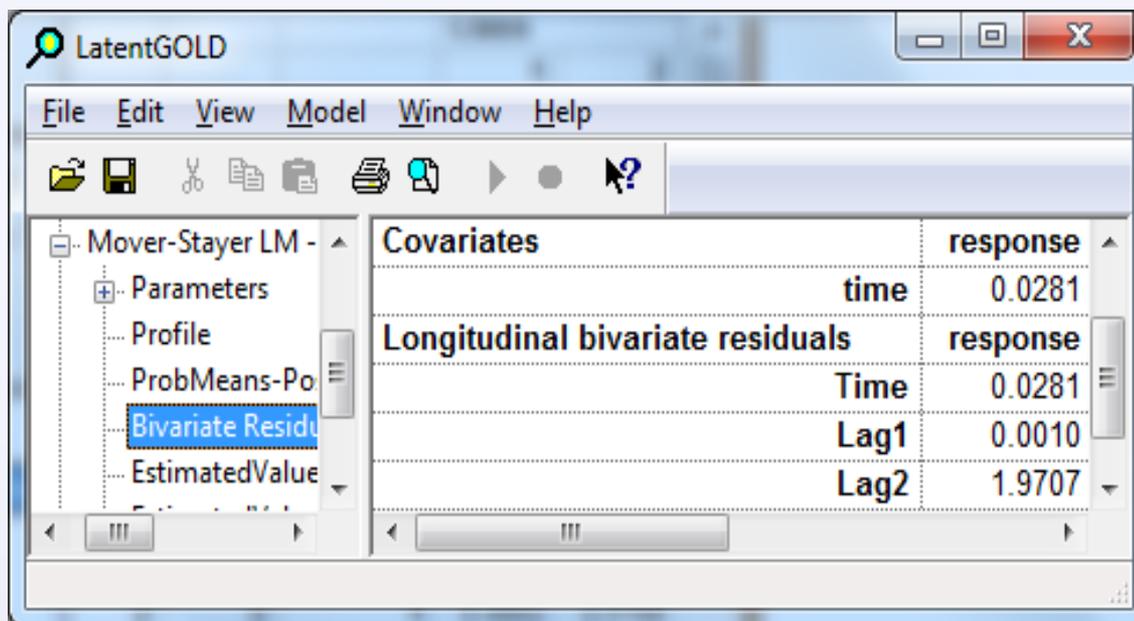
Stayer class showing 61.75% satisfied each year

Mover class showing changes over time

## Example 2: Longitudinal-Plot with Predicted Probability of Being Satisfied ('Overall Prob') Appended



## Example 2: L-BVRs for the Mover-Stayer Model



The screenshot shows the LatentGOLD software interface. The left pane displays the model structure for 'Mover-Stayer LM', with 'Bivariate Residuals' selected. The right pane shows the parameter estimates for these residuals.

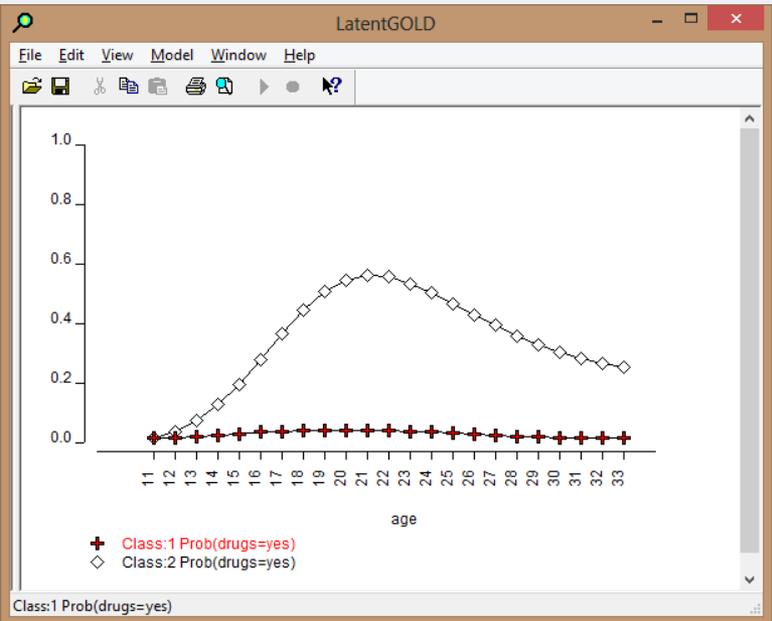
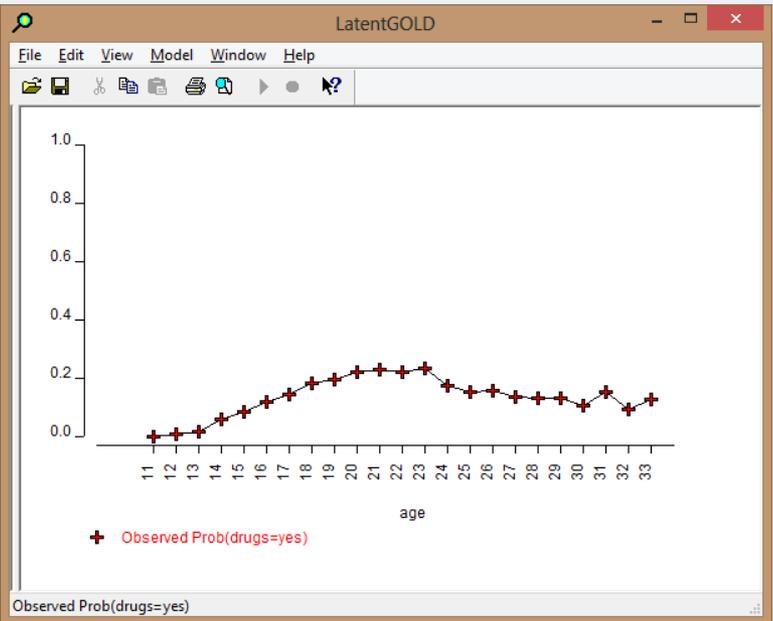
Covariates		response
	time	0.0281
Longitudinal bivariate residuals		response
	Time	0.0281
	Lag1	0.0010
	Lag2	1.9707

# Example 3: Latent GOLD Longitudinal Analysis of Sparse Data

	id	age	time	male	ethn4	drugs	time_2
1	1	11	0	1	1	.	0
2	1	12	1	1	1	.	3
3	1	13	2	1	1	0	6
4	1	14	3	1	1	0	9
5	1	15	4	1	1	0	12
6	1	16	5	1	1	0	15
7	1	17	6	1	1	1	18
8	1	18	7	1	1	.	21
9	1	19	8	1	1	.	24
10	1	20	9	1	1	1	27
11	1	21	10	1	1	.	30

- N=1725 pupils who were of age 11-17 at the initial measurement occasion (in 1976)
- Survey conducted annually from 1976 to 1980 and at three year intervals after 1980
- 23 time points (T+1=23), where t=0 corresponds to age 11 and the last time point to age 33.
- For each subject, data is observed for at most 9 time points (the average is 7.93) which means that responses for the other time points are treated as missing. (See Figure 2)
- Dichotomous dependent variable – ‘drugs’ indicating whether respondent used hard drugs during the past year (1=yes; 0=no).
- Time-varying predictors are ‘time’ (t) and ‘time\_2’ (t<sup>2</sup>); time-constant predictors are ‘male’ and ‘ethn4’ (ethnicity).

# Example 3: Latent GOLD Longitudinal Analysis of Sparse Data



The plot on the left shows the overall trend in drug usage during this period is non-linear, with zero usage reported for 11 year olds, increasing to a peak in the early 20s and then declining through age 33. The plot on the right plots the results from a mixture latent Markov model suggesting that the population consists of 2 distinct segments with different growth rates, Class 2 consisting primarily of non-users.

# Example 3: 2-class MLM Model

LatentGOLD

File Edit View Model Window Help

NYS9wave.sav

- null -  $L^2 = 4499.8072$
- 1-class regression -  $L^2 = 4621$
- 2-class regression -  $L^2 = 2935$
- model a: 2-state; age and ag
- model b: model a + 2 class -
- Parameters
- Profile
- Profile-Longitudinal
- ProbMeans-Posterior
- Bivariate Residuals
- EstimatedValues-Regress
- **EstimatedValues-Model**
- Model6

				Class	
				1	2
				0.6586	0.3414

				State[=0]	
				1	2
				0.9993	0.0007
				0.9990	0.0010

				State	
				1	2
time	time_2	Class	State[-1]	0.9971	0.0029
1	1	1	1	0.0496	0.9504
1	1	2	1	0.9712	0.0288
1	1	2	2	0.0012	0.9988
2	1	1	1	0.9954	0.0046
21	441	2	2	0.0889	0.9111
22	484	1	1	0.9993	0.0007
22	484	1	2	0.7453	0.2547
22	484	2	1	0.9930	0.0070
22	484	2	2	0.0643	0.9357

			drugs	
			no	yes
			0.9886	0.0114
			0.1379	0.8621

Class size

Initial state probabilities by class

Transition probabilities by class

Measurement model probabilities

# Example 3: Including Gender and Ethnicity as Covariates in Model

Markov - NYS9wave.sav - model b: model a + 2 class

Variables | Advanced | Model | Residuals | ClassPred | Output | Technical

age .  
psu1976 .  
ethn2 .  
ethn3 .  
alc .  
mrj .  
faslt .  
ftheft .

Indicators--> drugs Nominal 3

Case ID--> id 1725

<<--Covariates

time	Num-Fixed	23
time_2	Num-Fixed	23
male	Nominal	
ethn4	Nominal	

States  
2

Case Weight-->

Selection  
Selection Variable-->

Lexical Order

Scan Reset

Close Cancel Estimate Help

Markov - NYS9wave.sav - model b: model a + 2 class

Variables | Advanced | Model | Residuals | ClassPred | Output | Technical

Models for Indicators

	States	Classes
drugs	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Model for Initial States

	Initial States	Class Dependent
time	<input type="checkbox"/>	<input type="checkbox"/>
time_2	<input type="checkbox"/>	<input type="checkbox"/>
male	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ethn4	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Model for Transitions

	Transitions	Class Dependent
time	<input checked="" type="checkbox"/>	<input type="checkbox"/>
time_2	<input checked="" type="checkbox"/>	<input type="checkbox"/>
male	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ethn4	<input checked="" type="checkbox"/>	<input type="checkbox"/>

State Independent  
 Error Variances  
 Error Covariances

Model for Classes

	Classes
time	<input type="checkbox"/>
time_2	<input type="checkbox"/>
male	<input checked="" type="checkbox"/>
ethn4	<input checked="" type="checkbox"/>

Close Cancel Estimate Help

# Example 3: Including Gender and Ethnicity as Covariates in Model

LatentGOLD

File Edit View Model Window Help

NYS9wave.sav

			LL	BIC(LL)	Npar
· null - $L^2 = 4499.807$	· null	1-Class Regression	-5027.9818	10227.3822	23
· 1-class regression -	· 1-class regression	1-Class Regression	-5088.6953	10199.7495	3
· 2-class regression -	· 2-class regression	2-Class Regression	-4245.6077	8543.3863	7
· model a: 2-state; age	· model a: 2-state; age and age sq affecting transitions	2-State 1-Class	-4009.8956	8086.8680	9
· model b: model a +	· model b: model a + 2 class	2-State 2-Class	-3992.6040	8082.0967	13
	· model c: model b + gender and ethnicity affecting classes	2-State 2-Class	-3975.1827	8077.0660	17

Done.

Model Summary output, showing that adding gender and ethnicity improves the fit.

# Example 3: Including Gender and Ethnicity as Covariates in Model

LatentGOLD

File Edit View Model Window Help

VS9wave.sav

null -  $L^2 = 4499.8072$   
 1-class regression -  $L^2 = 462$   
 2-class regression -  $L^2 = 293$   
 model a: 2-state; age and ag  
 model b: model a + 2 class -  
 model c: model b + gender i

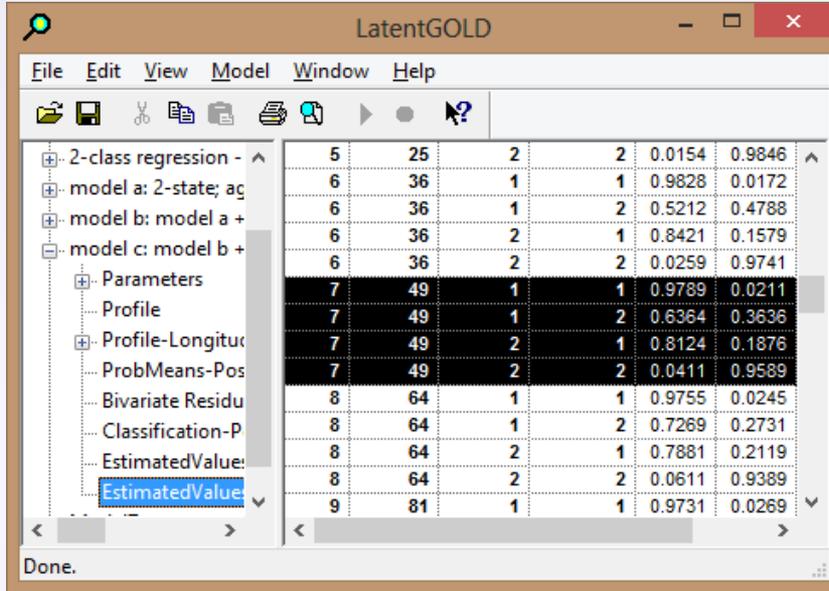
- Parameters
- Profile
- Profile-Longitudinal
- ProbMeans-Posterior
- Bivariate Residuals
- Classification-Posterior**
- EstimatedValues-Regress
- EstimatedValues-Model
- Iteration Detail

Model7

id	age	time	time_2	male	ethn4	drugs	ObsFreq	Class		State			
								Modal	1	2	Modal	1	2
1	11	0	0	male	white	.	1.0000	2	0.0002	0.9998	1	1.0000	0.0000
1	12	1	1	male	white	.	1.0000	2	0.0002	0.9998	1	0.9998	0.0002
1	13	2	4	male	white	no	1.0000	2	0.0002	0.9998	1	0.9997	0.0003
1	14	3	9	male	white	no	1.0000	2	0.0002	0.9998	1	0.9985	0.0015
1	15	4	16	male	white	no	1.0000	2	0.0002	0.9998	1	0.9869	0.0131
1	16	5	25	male	white	no	1.0000	2	0.0002	0.9998	1	0.8844	0.1156
1	17	6	36	male	white	yes	1.0000	2	0.0002	0.9998	2	0.0359	0.9641
1	18	7	49	male	white	.	1.0000	2	0.0002	0.9998	2	0.0412	0.9588
1	19	8	64	male	white	.	1.0000	2	0.0002	0.9998	2	0.0360	0.9640
1	20	9	81	male	white	yes	1.0000	2	0.0002	0.9998	2	0.0020	0.9980
1	21	10	100	male	white	.	1.0000	2	0.0002	0.9998	2	0.0608	0.9392
1	22	11	121	male	white	.	1.0000	2	0.0002	0.9998	2	0.0704	0.9296
1	23	12	144	male	white	yes	1.0000	2	0.0002	0.9998	2	0.0074	0.9926
1	24	13	169	male	white	.	1.0000	2	0.0002	0.9998	2	0.2465	0.7535
1	25	14	196	male	white	.	1.0000	2	0.0002	0.9998	2	0.4454	0.5546
1	26	15	225	male	white	no	1.0000	2	0.0002	0.9998	1	0.6282	0.3718
1	27	16	256	male	white	.	1.0000	2	0.0002	0.9998	2	0.3958	0.6042
1	28	17	289	male	white	.	1.0000	2	0.0002	0.9998	2	0.2034	0.7966
1	29	18	324	male	white	yes	1.0000	2	0.0002	0.9998	2	0.0435	0.9565
1	30	19	361	male	white	.	1.0000	2	0.0002	0.9998	2	0.1783	0.8217
1	31	20	400	male	white	.	1.0000	2	0.0002	0.9998	2	0.2688	0.7312
1	32	21	441	male	white	.	1.0000	2	0.0002	0.9998	2	0.3294	0.6706

Done.

# Example 3: Including Gender and Ethnicity as Covariates in Model



The screenshot shows the LatentGOLD software interface. The main window displays a table of transition probabilities. The table has 10 rows and 6 columns. The first two columns represent age and state, the next two represent class, and the last two represent transition probabilities. The table is as follows:

Age	State	Class	Class	Transition Probability 1	Transition Probability 2
5	25	2	2	0.0154	0.9846
6	36	1	1	0.9828	0.0172
6	36	1	2	0.5212	0.4788
6	36	2	1	0.8421	0.1579
6	36	2	2	0.0259	0.9741
7	49	1	1	0.9789	0.0211
7	49	1	2	0.6364	0.3636
7	49	2	1	0.8124	0.1876
7	49	2	2	0.0411	0.9589
8	64	1	1	0.9755	0.0245
8	64	1	2	0.7269	0.2731
8	64	2	1	0.7881	0.2119
8	64	2	2	0.0611	0.9389
9	81	1	1	0.9731	0.0269

The interface also shows a menu bar (File, Edit, View, Model, Window, Help) and a toolbar with various icons. The left sidebar contains a tree view of the model structure, with 'EstimatedValues' selected. The status bar at the bottom says 'Done.'

For concreteness we will focus on 18 year olds (see highlighted cells in figure). We see that 18 year olds who were in the lower usage state (State 1) at age 17 have a probability of .1876 of switching to the higher usage state (State 2) if they are in Class 2 compared to a probability of only .0211 of switching if they were in Class 1. In addition, if they were in the higher use state (State 2) at age 17, they have a probability of .9589 of remaining in that state compared to only .3636 if they were in Class 1. Thus, based on these different transition probabilities we see that Class 2 is more likely to move to and remain in a higher drug usage state than Class 1.

## Appendix: Equations

Latent Markov:

$$P(\mathbf{y}_i) = \sum_{x_0=1}^K \sum_{x_1=1}^K \dots \sum_{x_T=1}^K P(x_0, x_1, \dots, x_T) P(\mathbf{y}_i | x_0, x_1, \dots, x_T)$$

$$P(x_0, x_1, \dots, x_T) = P(x_0) \prod_{t=1}^T P(x_t | x_{t-1})$$

$$P(\mathbf{y}_i | x_0, x_1, \dots, x_T) = \prod_{t=0}^T P(\mathbf{y}_{it} | x_t) = \prod_{t=0}^T \prod_{j=1}^J P(y_{itj} | x_t)$$

Latent Growth:

$$P(\mathbf{y}_i | \mathbf{z}_i) = \sum_{w=1}^L P(w | \mathbf{z}_i) \prod_{t=0}^T P(y_{it} | w, \mathbf{z}_{it})$$

# References

Vermunt, J.K., Tran, B. and Magidson, J (2008). Latent Markov model.

Vermunt, J.K. Langeheine, R., Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.