

# Latent class analysis for intensive longitudinal data, Hidden Markov processes, Regime switching models and Dynamic Structural Equations in Mplus

Tihomir Asparouhov, Bengt Muthén and Ellen Hamaker

May 24, 2016

- Motivation
- Dynamic Structural Equations Model (DSEM) framework and estimation
- New Multilevel Mixture Models: these are needed as building block for the more advanced models
- Single level models: HMM (Hidden Markov Models), MSAR (Markov Switching Auto-Regressive), MSKF (Markov Switching Kalman Filter)
- Two-level HMM, MSAR, MSKF

- Merge "time series", "structural equation", "multilevel" and "mixture" modeling concepts in a generalized modeling framework in Mplus V8
- In this context two-level means single-level. Cluster is always the individual. Many observations are collected within subject and analyzed in long format. Most time-series models were developed for single level data however most social science applications need two-level methods because we study many individuals across time, rather than the US economy across time.
- Mplus release timeline: V8 will have DSEM and probably single level MSAR. V8.1 will have two-level MSAR.

- Consider the following hypothetical example. A group of depression patients answer daily a brief survey to evaluate their current state. Based on current observations, past history, most recent history, similar behavior from other patients we classify the patient in one of 3 states:
  - S1: OK
  - S2: Stretch of poor outcomes, needs doctor visit/evaluation
  - S3: At risk for suicide, needs hospitalization
- Future of health care? Cheaper, smarter and more effective?
- The models we describe in this talk can be used to model the data from this hypothetical example: combine mixture, multilevel, time-series, latent variables, and structural models.

- The goal of these models is to provide continuous time monitoring for unobserved categorical and continuous latent constructs. This is needed to study latent variable development across time and to be able to detect problems earlier and prevent/react.
- Modeling two distinct sources of correlation: within individual correlations can be due to subject-specific effect (two-level modeling) or it can be due to correlation due to proximity of observations (autocorrelation) of the observations (time series modeling). The two types of correlations are easy to parse out from the data in sufficiently long longitudinal data.
- Various mobile devices are now utilized for collection of data: ecological momentary assessment (EMA) data and experience sampling methods (ESM). Increased need for time intensive methods.

- For more DSEM information see about 200 slides from the pre-conference workshop. Here is a quick review.
- We use a simplified DSEM version that excludes TVEM (no cross-classified modeling)
- Let  $Y_{it}$  be an observed vector of measurements for individual  $i$  at time  $t$ .
- Let  $\eta_{it}$  be a latent variable vector for individual  $i$  at time  $t$ .
- Let  $X_{it}$  be a observed vector of covariates individual  $i$  at time  $t$ .
- Similarly  $Y_i$ ,  $X_i$  and  $\eta_i$  are individual-specific variables, time invariant
- Main decomposition equation

$$Y_{it} = Y_{1,it} + Y_{2,i}$$

- $Y_{2,i}$  are the "individual" specific contribution, normal latent variable.  $Y_{1,it}$  is the residual.

- The within level model includes previous periods latent variables  $\eta_{i,t-l}$  as predictors, modeling the correlations in consecutive periods.
- $L$  is the lag variable: how many time period back are included in the model.

$$Y_{1,it} = \nu_1 + \sum_{l=0}^L \Lambda_{1,l} \eta_{i,t-l} + \varepsilon_{it}$$

$$\eta_{i,t} = \alpha_1 + \sum_{l=0}^L B_{1,l} \eta_{i,t-l} + \Gamma_1 x_{it} + \xi_{it}.$$

- The usual structural equations at level 2. Every within level parameter can be random effect: part of  $\eta_i$

$$Y_{2,i} = \nu_2 + \Lambda_2 \eta_i + \varepsilon_i$$
$$\eta_i = \alpha_2 + B_2 \eta_i + \Gamma_2 x_i + \xi_i$$

- Observed variables can also have lag variables and can be used as predictors.
- Ordered polytomous and binary dependent variables are included in this framework using the underlying  $Y^*$  approach: probit link
- The model is also of interest when  $N=1$ . No second level. All observations are correlated. Multivariate econometrics models.
- The  $N = 1$  model can be used also on the data from a single individual to construct a psychological profile and match it to the known profile of a psychological disorder.
- We use Bayes estimation
- The above model has variables with negative or zero indices. We treat those as auxiliary parameters that have prior distribution. Automatic option specification is implemented in Mplus.



- Let  $S_{it}$  be a categorical latent variable for individual  $i$  at time  $t$ . We call it "State". In Mplus this is a within level latent variable.  $S$  can take values  $1, 2, \dots, K$  where  $K$  is the number of classes/states in the model.
- The model on the between level is not affected by  $S$ . The model on the within level is state specific

$$[Y_{1,it}|S_{it} = s] = \nu_{1,s} + \sum_{l=0}^L \Lambda_{1,l,s} \eta_{i,t-l} + \varepsilon_{it}$$

$$[\eta_{i,t}|S_{it} = s] = \alpha_{1,s} + \sum_{l=0}^L B_{1,l,s} \eta_{i,t-l} + \Gamma_{1,s} x_{it} + \xi_{it}$$

- The residual variance  $Var(\varepsilon_{it}|S_{it} = s)$  and  $Var(\xi_{it}|S_{it} = s)$  are also state specific

# DSEM Mixture Model Continued

- In addition to the above model we have to specify the distribution of  $S_{it}$

$$P(S_{it} = s) = \frac{\text{Exp}(\alpha_{is})}{\sum_{s=1}^K \text{Exp}(\alpha_{is})}$$

- $\alpha_{is}$  are normally distributed random effects, i.e., they are part of the vector  $\eta_i$ . For identification the last one is zero  $\alpha_{iK} = 0$
- Individual level predictors can be used to predict  $\alpha_{is}$ , regress the logits on covariates.
- In MCMC we use Metropolis Hastings to update the random effects. Using the proposal distribution  $N(\alpha_{is}, \Sigma)$  where  $\Sigma$  is the model estimate for  $\Sigma = \text{Var}(\alpha_{is})$  from a burnin period. The new draw  $\hat{\alpha}_{is}$  is accepted with probability

$$\text{Acceptance ratio} = \frac{\text{Prior}(\hat{\alpha}_{is}) \text{Likelihood}(S_{it} | \hat{\alpha}_{is})}{\text{Prior}(\alpha_{is}) \text{Likelihood}(S_{it} | \alpha_{is})}$$

- Potential avenues to improve mixing if needed
  - Use  $c\Sigma$  where  $c$  is a constant regulating the acceptance rate within a desirable range of 15% to 35%
  - Use cluster specific proposal distribution from a burnin period  $c\Sigma_i$  where  $\Sigma_i = \text{Var}(\alpha_{is} | \text{data})$ . Unbalanced designs most likely will need this.
  - Separate  $\alpha_{is}$  from other random effects.
- Limited simulations show good performance even without these added steps, however, final version will include these probably.

# Bayes Multilevel Mixture Model

- Bayes Multilevel Mixture Model = DSEM Mixture without the lag variables  $\eta_{i,t-l}$ . For the next 20 slides no intensive longitudinal data.
- The model is essentially the Twolevel Mixture model implemented in Mplus which can be estimated with ML.
- Advantage of this model is that we have Bayes estimation and thus can estimate models with any number of random effects.
- With ML,  $\alpha_{is}$  typically can not be all estimated and we constrain them to be proportional via a factor to reduce the number of numerical integration from  $K - 1$  to 1. With Bayes we don't need to do that.
- Asparouhov, T. & Muthen, B. (2008). Multilevel mixture models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), *Advances in latent variable mixture models*, pp. 27-51. Charlotte, NC: Information Age Publishing, Inc.

Most common approach approach for LCA with clustering (nested data) is to use ML single level with robust/sandwich SE that take the clustering into account. Three problems

- Does not allow cluster specific class distribution.
- Assumes full measurement invariance for the latent class variable measurement model
- If these assumptions are not met likely to yield spurious classes

- For indicator  $p$  individual  $i$  in cluster  $j$

$$P(U_{pij} = 1 | C_{ij} = k) = \Phi(\tau_{pk} + \tau_{pj})$$

$$P(C_{ij} = k) = \frac{\text{Exp}(\alpha_j + \alpha_{jk})}{\sum_{s=1}^K \text{Exp}(\alpha_j + \alpha_{js})}$$

- $\tau_{pk}$  is a non-random parameter (the usual threshold parameter)
- $\tau_{pj}$  is a measurement non-invariance zero mean random effect that allows certain indicators to be more or less frequent in cluster  $j$  than the population values, beyond what the latent class distribution explains. For example, certain measurement instruments not universally accurate.
- $\alpha_j$  are non-random effects that fits the population level class distribution
- $\alpha_{jk}$  are zero mean random effects that allow cluster specific class distribution
- ML estimation would use 10 dim integration. Bayes 40 sec rep.

# LCA with measurement non-invariance: 3 Class simulation

```
MONTECARLO:  NAMES ARE u1-u8;
              NOBS = 5000;
              NREP = 100;
              NCSIZES = 1;
              CSIZES = 100(50);
              classes=c(3);
              genclasses=c(3);
              generate=u1-u8(1);
              categorical=u1-u8;

ANALYSIS:    TYPE IS mixture twolevel;
              estimator = bayes; proc=2;

MODEL MONTECARLO:

%WITHIN%
%overall%

%BETWEEN%
%overall%
C#1-C#2*0.3;
[C#1*0.8 C#2*0.4];
u1-u8*0.2;

%C#1%
[u1$1-u8$1*1];

%C#2%
[u1$1-u8$1*-1];

%C#3%
[u1$1-u4$1*-1];
[u5$1-u8$1*1];
```

# LCA with measurement non-invariance: Simulation results

## Latent Class 3

### Thresholds

U1\$1	-1.000	-0.9922	0.0788	0.0816	0.0062	0.970	1.000
U2\$1	-1.000	-1.0027	0.0725	0.0814	0.0052	0.990	1.000
U3\$1	-1.000	-0.9994	0.0791	0.0816	0.0062	0.950	1.000
U4\$1	-1.000	-1.0027	0.0805	0.0810	0.0064	0.980	1.000
U5\$1	1.000	1.0124	0.0772	0.0812	0.0061	0.960	1.000
U6\$1	1.000	1.0047	0.0872	0.0801	0.0075	0.930	1.000
U7\$1	1.000	1.0114	0.0748	0.0808	0.0057	0.960	1.000
U8\$1	1.000	1.0035	0.0797	0.0802	0.0063	0.960	1.000

### Variances

U1	0.200	0.2049	0.0354	0.0405	0.0013	0.980	1.000
U2	0.200	0.2074	0.0396	0.0411	0.0016	0.970	1.000
U3	0.200	0.2064	0.0330	0.0406	0.0011	0.970	1.000
U4	0.200	0.2125	0.0395	0.0417	0.0017	0.930	1.000
U5	0.200	0.2108	0.0383	0.0413	0.0016	0.960	1.000
U6	0.200	0.2086	0.0332	0.0409	0.0012	0.980	1.000
U7	0.200	0.2129	0.0360	0.0415	0.0015	0.970	1.000
U8	0.200	0.2136	0.0425	0.0418	0.0020	0.910	1.000

## Between Level

### Means

C#1	0.800	0.8120	0.0805	0.0789	0.0066	0.950	1.000
C#2	0.400	0.4057	0.0874	0.0832	0.0076	0.930	1.000

### Variances

C#1	0.300	0.3151	0.0790	0.0763	0.0064	0.910	1.000
C#2	0.300	0.3336	0.0780	0.0839	0.0072	0.940	1.000



# Unrestricted Two-Level Mixture Model Simulation

- $Y_{ij}$  is a vector of observed continuous variables for individual  $i$  in cluster  $j$

$$Y_{ij} = Y_{b,j} + Y_{w,ij}$$

$$[Y_{w,ij} | C_{ij} = k] \sim N(\mu_k, \Sigma_k)$$

$$P(C_{ij} = k) = \frac{\text{Exp}(\alpha_j + \alpha_{jk})}{\sum_{s=1}^K \text{Exp}(\alpha_j + \alpha_{js})}$$

- This model requires numerical integration in Mplus even if the latent class variable is observed
- With Bayes it can even accommodate categorical variables
- It makes feasible the models discussed in Asparouhov, T. & Muthén, B. (2012). Multiple group multilevel analysis. Mplus Web Notes: No. 16.

# Unrestricted Two-Level Mixture Model Simulation

MODEL:

```
%WITHIN%  
%overall%
```

```
%C#1%  
[y1-y3*-1]; y1-y3*1;  
y1 with y2*0.2; y1 with y3*0.3; y2 with y3*0.3;
```

```
%C#2%  
[y1-y3*1]; y1-y3*0.6;  
y1 with y2*0.3; y1 with y3*0.4; y2 with y3*0.2;
```

```
%BETWEEN%  
%overall%  
C#1*0.5; [C#1*0.8];  
y1-y3*1; y1 with y2*0.3; y1 with y3*0.4; y2 with y3*0.1;
```

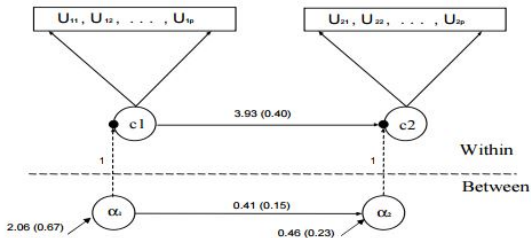
# Unrestricted Two-Level Mixture Model Simulation Results

Within Level								
Latent Class 1								
Y1	WITH							
Y2		0.200	0.1997	0.0167	0.0160	0.0003	0.930	1.000
Y3		0.300	0.2983	0.0157	0.0163	0.0002	0.970	1.000
Means								
Y1		-1.000	-1.0053	0.0728	0.0653	0.0053	0.930	1.000
Y2		-1.000	-1.0019	0.0777	0.0666	0.0060	0.920	1.000
Variances								
Y1		1.000	1.0035	0.0203	0.0194	0.0004	0.930	1.000
Y2		1.000	1.0029	0.0215	0.0205	0.0005	0.930	1.000
Y3		1.000	1.0016	0.0193	0.0202	0.0004	0.940	1.000
Between Level								
Y1	WITH							
Y2		0.300	0.3083	0.0802	0.0801	0.0064	0.970	1.000
Y3		0.400	0.4157	0.0822	0.0831	0.0069	0.910	1.000
Variances								
Y1		1.000	1.0203	0.1150	0.1078	0.0135	0.930	1.000
Y2		1.000	1.0286	0.0919	0.1085	0.0092	0.980	1.000
Categorical Latent Variables								
Means								
C#1		0.800	0.8052	0.0620	0.0619	0.0038	0.960	1.000
Variances								
C#1		0.500	0.5150	0.0740	0.0726	0.0056	0.920	1.000

# Multilevel Latent Transition Analysis (MLTA) with cluster specific transition probabilities

- Hypothetical example: Students are nested within schools and are classified in 2 classes at two separate occasions. We are interested in how the  $P(C_2|C_1)$  varies across schools.

Figure 1: Two-level Latent Transition Model



# Multilevel Latent Transition Analysis (MLTA) with cluster specific transition probabilities

- The model that can be estimated in Mplus with ML

$$P(C_{1ij} = c) = \frac{\exp(\alpha_{1cj} + \beta_{1cj}x_{1ij})}{\sum_c \exp(\alpha_{1cj} + \beta_{1cj}x_{1ij})}$$

$$P(C_{2ij} = d | C_{1ij} = c) = \frac{\exp(\alpha_{2dj} + \gamma_{dcj} + \beta_{2dj}x_{2ij})}{\sum_d \exp(\alpha_{2dj} + \gamma_{dcj} + \beta_{2dj}x_{2ij})}$$

- However  $\gamma_{dcj}$  does not really vary across clusters it is really  $\gamma_{dc}$
- Even if we regress  $\alpha_{2j}$  on  $\alpha_{1j}$  (equivalent to correlation) we still have just 2 random effects for the joint distribution of two binary latent class variables while the degrees of freedom is 3
- Current Mplus ML estimation  $P(C_2|C_1)$  does not fully vary

# Multilevel Latent Transition Analysis (MLTA) with cluster specific transition probabilities

- New Bayes model

$$P(C_{1,ij} = k_1) = \frac{\text{Exp}(\alpha_{jk_1})}{\sum_{k_1=1}^K \text{Exp}(\alpha_{jk_1})}$$

$$P(C_{2,ij} = k_2 | C_{1,ij} = k_1) = \frac{\text{Exp}(\alpha_{jk_1k_2})}{\sum_{k_2=1}^K \text{Exp}(\alpha_{jk_1k_2})}$$

- The old model is equivalent to the assumption that  $\alpha_{j21} - \alpha_{j11}$  is the constant  $\gamma$
- $\alpha_{jk_1K} = 0$  for identification
- Note that the transition probabilities can also be regressed on predictors by regressing  $\alpha_{jk_1k_2}$  on predictors

- A model with two binary class variables has 3 random effects (3 df in joint  $C_1$  by  $C_2$  disytribution

$$\alpha_{j1} = \log(P(C_1 = 1)/P(C_1 = 2))$$

$$\alpha_{j11} = \log(P(C_2 = 1|C_1 = 1)/P(C_2 = 2|C_1 = 1))$$

$$\alpha_{j21} = \log(P(C_2 = 1|C_1 = 2)/P(C_2 = 2|C_1 = 2))$$

- In MCMC the update for  $\alpha_{jk_1k_2}$  are no different than  $\alpha_{jk_1}$ , i.e, we use MH
- Simulation studies show that clusters sizes should not be small ( $> 50$ ). If cluster sizes are small joint tables will have empty cells that lead to logits of infinity. The results of that is biased overestimation for  $Var(\alpha_{jk_1k_2})$ .

```
montecarlo:
  names are u11-u14 u21-u24;
  genclasses = c1(2) c2(2);
  classes = c1(2) c2(2);
  nobs = 5000;
  ncsizes = 1;
  csizes = 100(50);
  nrep = 100;

ANALYSIS:  TYPE = MIXTURE TWOLEVEL RANDOM;
           ESTIMATOR = BAYES;

MODEL population:

    %within%
    %OVERALL%
    s1 | C2#1 on C1#1;
    s2 | C2#1 on C1#2;
    u11-u14*1; u21-u24*1;

    %BETWEEN%
    %OVERALL%
    [C1#1*0.5]; C1#1*0.05;
    s1-s2*0.05; [s1*-0.5 s2*1];
    u11-u14*1; u21-u24*1;
```



```
MODEL POPULATION-C1:
```

```
    %within%  
    %c1#1%  
    [u11-u14*1.0];  
    %c1#2%  
    [u11-u14*-1.0];
```

```
MODEL POPULATION-C2:
```

```
    %within%  
    %c2#1%  
    [u21-u24*1.0];  
    %c2#2%  
    [u21-u24*-1.0];
```

# MLTA Simulation Results

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	¶ Sig Coeff
Within Level							
Latent Class Pattern 1 1							
Means							
U11	1.000	1.0071	0.0929	0.1010	0.0086	0.930	1.000
U24	1.000	0.9959	0.0976	0.1023	0.0094	0.970	1.000
Variances							
U11	1.000	1.0009	0.0220	0.0216	0.0005	0.970	1.000
U24	1.000	0.9995	0.0232	0.0216	0.0005	0.920	1.000
Latent Class Pattern 1 2							
Means							
U11	1.000	1.0071	0.0929	0.1010	0.0086	0.930	1.000
U24	-1.000	-1.0109	0.0963	0.1024	0.0093	0.950	1.000
Variances							
U11	1.000	1.0009	0.0220	0.0216	0.0005	0.970	1.000
U24	1.000	0.9995	0.0232	0.0216	0.0005	0.920	1.000

# MLTA Simulation Results

## Between Level

### Means

S1	-0.500	-0.5007	0.0492	0.0479	0.0024	0.940	1.000
S2	1.000	1.0074	0.0602	0.0648	0.0036	0.990	1.000

### Variances

U11	1.000	1.0438	0.1492	0.1583	0.0240	0.940	1.000
U24	1.000	1.0459	0.1586	0.1584	0.0270	0.910	1.000
S1	0.050	0.0691	0.0274	0.0317	0.0011	0.910	1.000
S2	0.050	0.0949	0.0533	0.0537	0.0048	0.890	1.000

## Categorical Latent Variables

### Between Level

### Means

C1#1	0.500	0.4973	0.0404	0.0390	0.0016	0.950	1.000
------	-------	--------	--------	--------	--------	-------	-------

### Variances

C1#1	0.050	0.0544	0.0237	0.0221	0.0006	0.900	1.000
------	-------	--------	--------	--------	--------	-------	-------

# Single Level LTA with Probability Parameterization

- In single level the logits of transition probabilities are not random effects. They are non-random parameters.
- Mplus has 3 different parameterizations for ML estimation of LTA: logit, loglinear, probability
- New Bayes estimation for LTA with probability parameterization: it allows for Lag=1 or Lag=2,  $P(C_2|C_1)$  or  $P(C_3|C_1, C_2)$  just like ML
- The model parameters are the probabilities directly  $P(C_1)$  and  $P(C_2|C_1)$  and  $P(C_3|C_1, C_2)$
- Easy MCMC implementation.  $P(C_1)$  and  $P(C_2|C_1)$  and  $P(C_3|C_1, C_2)$  have conjugate Dirichlet prior.

# LTA with Probability Parameterization Simulation

## MONTECARLO:

```
NAMES = u11-u16 u21-u26;  
GENERATE = u11-u16 u21-u26(1);  
CATEGORICAL = u11-u16 u21-u26;  
GENCLASSES = c1(3) c2(3);  
CLASSES = c1(3) c2(3);  
NOBSERVATIONS = 1000; NREPS = 100;
```

## ANALYSIS:

```
TYPE = MIXTURE; estimator=bayes;  
PARAMETERIZATION = PROBABILITY; proc=2
```

## MODEL POPULATION:

```
%OVERALL%  
[c1#1*0.3 c1#2*0.2];  
c2#1 ON c1#1*0.7; c2#2 ON c1#1*0.2;  
c2#1 ON c1#2*0.1; c2#2 ON c1#2*0.7;  
c2#1 ON c1#3*0.2; c2#2 ON c1#3*0.2;
```

## MODEL POPULATION-c1:

```
%c1#1%  
[u11$1-u16$1*1];  
%c1#2%  
[u11$1-u13$1*1];  
[u14$1-u16$1*-1];  
%c1#3%  
[u11$1-u16$1*-1];
```

# LTA with Probability Parameterization Results

## Categorical Latent Variables

C2#1	ON							
C1#1		0.700	0.6962	0.0315	0.0332	0.0010	0.980	1.000
C1#2		0.100	0.0978	0.0367	0.0341	0.0013	0.960	1.000
C1#3		0.200	0.1996	0.0221	0.0216	0.0005	0.930	1.000
C2#2	ON							
C1#1		0.200	0.2027	0.0338	0.0320	0.0011	0.920	1.000
C1#2		0.700	0.6958	0.0599	0.0489	0.0036	0.910	1.000
C1#3		0.200	0.1990	0.0268	0.0249	0.0007	0.940	1.000
Means								
C1#1		0.300	0.3001	0.0155	0.0166	0.0002	0.960	1.000
C1#2		0.200	0.2049	0.0173	0.0178	0.0003	0.960	1.000

# Single Level Hidden Markov Models

- Data is again intensive longitudinal. First we consider the single level model,  $N=1$ . We discuss 3 models
  - Hidden Markov Model (HMM)
  - Markov Switching Autoregressive (MSAR)
  - Markov Switching Kalman Filter (MSKF)
- DSEM allows for time series modeling for observed and latent continuous variables. DSEM Mixture does not allow auto correlation for the latent categorical variable
- In time series data it is not realistic to assume that  $S_t$  and  $S_{t-1}$  are independent, where  $S_t$  is the state/class variable at time  $t$ . On the contrary. A realistic model will allow  $S_t$  and  $S_{t-1}$  be highly correlated if the observations are taken very frequently.

# Hidden Markov Models

- The model has two parts: measurement part and Markov switching part
- The measurement part is like any other Mixture model, it is defined by  $P(Y_t|S_t)$  where  $Y_t$  is a vector of observed variables and  $S_t$  is the latent class/state variable at time  $t$
- The Markov switching (regime switching) part is given by  $P(S_t|S_{t-1})$ . We use the same probability parametrization based on Dirichlet conjugate priors that we used with two latent class variables. The transition model  $Q = P(S_t|S_{t-1})$  has  $K(K-1)$  probability parameters. The transition matrix is  $K$  by  $K$  but the columns add up to 1



- Note that  $p=P(S_t)$  is not a model parameter. The probability is implicit and is the distribution of  $S_t$  if the sequence is observed to infinity. It can be obtained implicitly from the stationarity assumption that  $P(S_t)$  is independent of  $t$ . i.e., from the equation  $Qp = p$ . Since the first  $K - 1$  equations added up give the last we need to replace the last equation with  $\text{Sum}(p)=1$  to solve it.
- The MCMC step that updates the latent class variable

$$P(C_t = k | C_{t-1}, C_{t+1}) = \frac{P(C_{t+1} | C_t = k)P(C_t = k | C_{t-1})}{\sum_{k=1}^K P(C_{t+1} | C_t = k)P(C_t = k | C_{t-1})}$$

$$P(C_t = k | C_{t-1}, C_{t+1}, Y_t) = \frac{P(Y_t | C_t = k)P(C_t = k | C_{t-1}, C_{t+1})}{\sum_{k=1}^K P(Y_t | C_t = k)P(C_t = k | C_{t-1}, C_{t+1})}$$

- $C_{t=0}$  is treated as an auxiliary parameter
- $C_{t=0}$  can be given a prior
- Mplus provides an automatic prior option. The prior is updated in the first 100 MCMC iteration which are consequently discarded and the prior is set to be the current sample distribution  $C_t$ . This is the default.
- If the length of the time series is long enough that prior does not matter

```
montecarlo:
  names are u1-u3;
  genclasses = c(2 &1);
  classes = c(2 &1);
  generate = u1-u3(1);
  categorical = u1-u3;
  nobs = 1000; nrep = 100;

analysis:
  type = mixture; proc=2; estimator = bayes;

model population:
  %overall%
  c#1 on c&1#1*0.90; c#1 on c&1#2*0.25;

model population-c:
  %c#1%
  [u1$1-u3$1*-1];
  %c#2%
  [u1$1-u3$1*1];
```

# HMM Results

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
Population		Average	Std. Dev.	Average		Cover	Coeff
Latent Class 1							
Thresholds							
U1\$1	-1.000	-0.9954	0.0642	0.0627	0.0041	0.950	1.000
U2\$1	-1.000	-1.0012	0.0664	0.0646	0.0044	0.930	1.000
U3\$1	-1.000	-0.9946	0.0627	0.0634	0.0039	0.920	1.000
Latent Class 2							
Thresholds							
U1\$1	1.000	1.0041	0.1117	0.1122	0.0124	0.910	1.000
U2\$1	1.000	1.0034	0.1087	0.1104	0.0117	0.940	1.000
U3\$1	1.000	1.0098	0.1245	0.1129	0.0154	0.930	1.000
Categorical Latent Variables							
C#1	ON						
C&1#1		0.900	0.8976	0.0132	0.0135	0.0002	0.940 1.000
C&1#2		0.250	0.2555	0.0261	0.0297	0.0007	0.970 1.000

- By combining HMM and DSEM we obtain a general model that includes time series for the latent class variable as well as factors and observed variables
- Markov Switching Autoregressive (MSAR) is simply the combination of Mixture-AR and HMM
- Markov Switching Kalman Filter (MSKF) is simply the combination of Mixture-Kalman Filter and HMM

- Let  $Y_{1t}$  and  $Y_{2t}$  be two observed variables at time  $t$  and  $S_t$  be the latent state variable at time  $t$  taking one of two states.

$$Y_{1t} = \alpha_{1,S_t} + \beta_{1,S_t} Y_{1,t-1} + \beta_{2,S_t} Y_{2,t-1} + \varepsilon_{1it}$$

$$Y_{2t} = \alpha_{2,S_t} + \beta_{3,S_t} Y_{1,t-1} + \beta_{4,S_t} Y_{2,t-1} + \varepsilon_{2it}$$

- The model has 20 parameters: 4  $\alpha$ , 8  $\beta$ , 6 residual covariance, and 2 parameters in the transition matrix:  $P(S_t = 1 | S_{t-1} = 1)$  and  $P(S_t = 1 | S_{t-1} = 2)$
- The model addresses the chicken or the egg problem. Example: wife and a husband mood in good and bad state.
- Hamaker, Grasman and Kamphuis (2016) Modeling BAS Dysregulation in Bipolar Disorder Illustrating the Potential of Time Series Analysis, Assessment. They argue that bipolar individuals show two state behaviour while controls don't.
- Simulation study speed of computation: it takes 1/3 of 1 second to estimate the model on a sample with 1000 observations

# Bivariate MSAR Simulation

montecarlo:

```
names are y1-y2;  
lagvar=y1-y2(1);  
genclasses = c(2 &1);  
classes = c(2 &1);  
nobs = 1000; nrep = 100;
```

analysis:

```
type = mixture; proc=2;  
estimator = bayes; parameterization=probability;
```

model population:

```
%overall%  
C#1 on C&1#1*0.7; C#1 on C&1#2*0.2;  
y1 y2 on y1&1*0 y2&1*0;
```

model population-c:

```
%c#1%  
[y1-y2*0]; y1-y2*1; y1 with y2*0.3;  
y1 on y1&1*0.4 y2&1*0.2;  
y2 on y2&1*0.5 y1&1*0.1;  
%c#2%  
[y1-y2*1.8]; y1-y2*1.3; y1 with y2*0.5;  
y1 on y1&1*0.6 y2&1*0;  
y2 on y2&1*0.3 y1&1*-0.2;
```

# Bivariate MSAR Results

Latent Class 2								
Y1	ON							
Y1&1		0.600	0.6057	0.0326	0.0345	0.0011	0.960	1.000
Y2&1		0.000	0.0022	0.0463	0.0455	0.0021	0.970	0.030
Y2	ON							
Y2&1		0.300	0.3027	0.0405	0.0425	0.0016	0.960	1.000
Y1&1		-0.200	-0.1978	0.0323	0.0329	0.0010	0.970	1.000
Y1	WITH							
Y2		0.500	0.5290	0.0734	0.0773	0.0062	0.930	1.000
Intercepts								
Y1		1.800	1.7482	0.1829	0.1624	0.0358	0.890	1.000
Y2		1.800	1.7738	0.1608	0.1454	0.0263	0.920	1.000
Residual Variances								
Y1		1.300	1.3477	0.1271	0.1193	0.0183	0.910	1.000
Y2		1.300	1.3207	0.0749	0.0888	0.0060	0.980	1.000
Categorical Latent Variables								
C#1	ON							
C&1#1		0.700	0.6973	0.0389	0.0429	0.0015	0.970	1.000
C&1#2		0.200	0.1933	0.0296	0.0297	0.0009	0.950	1.000



# Markov Switching Kalman Filter (MSKF)

- Three factor indicators  $Y_{jt}$  measuring a factor  $\eta_t$ .  $S_t$  is a two state categorical latent variable.
- We estimate a hidden Markov model for  $S_t$ , i.e.,  $P(S_t = 1|S_{t-1} = 1)$  and  $P(S_t = 1|S_{t-1} = 2)$  are probability parameters independent of  $t$ .
- For  $j=1,2,3$

$$Y_{jt} = v_j + \lambda_j \eta_t + \varepsilon_{jt}$$

$$\eta_t = \alpha_{S_t} + \beta_{1,S_t} \eta_{t-1} + \beta_{2,S_t} \eta_{t-2} + \xi_t$$

- MSAR(2) model for the factor
- For identification purposes  $\alpha_1 = 0$  and  $\lambda_1 = 1$  (this sets the scale of the latent variable to be the same as the first indicator, which is probably a better parameterization than fixing the variance of the factor to 1?)

```
montecarlo:
  names are y1-y3;
  genclclasses = c(2 &1);
  classes = c(2 &1);
  nobs = 2000; nrep = 100;

analysis:
  type = mixture; proc=2; biter=(5000);
  estimator = bayes; parameterization=probability;

model population:

  %overall%
  C#1 on C&1#1*0.7;
  C#1 on C&1#2*0.2;
  y1-y3*1; [y1-y3*0] (m1-m3);
  e by y1@1 y2-y3*1 (&2); e@1;
  e on e&1*0.3 e&2*0.2;

model population-c:
  %c#1%
  [e@0]; y1-y2*1; e*1;
  e on e&1*0.3 e&2*0.2;
  %c#2%
  [e*2]; y1-y2*1.2; e*1.2;
  e on e&1*0.6 e&2*0;
```

# MSKF Results

## Latent Class 2

E	BY							
Y1		1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Y2		1.000	1.0008	0.0188	0.0170	0.0003	0.900	1.000
Y3		1.000	0.9973	0.0167	0.0166	0.0003	0.970	1.000
E	ON							
E&1		0.600	0.6172	0.0496	0.0453	0.0027	0.910	1.000
E&2		0.000	-0.0094	0.0363	0.0369	0.0014	0.960	0.040
Intercepts								
Y1		0.000	-0.0754	0.2849	0.2223	0.0860	0.860	0.140
Y2		0.000	-0.0705	0.2809	0.2236	0.0831	0.870	0.130
Y3		0.000	-0.0608	0.2867	0.2227	0.0850	0.870	0.130
E		2.000	1.9463	0.1597	0.1575	0.0281	0.950	1.000
Residual Variances								
Y1		1.200	1.2045	0.0746	0.0738	0.0055	0.940	1.000
Y2		1.200	1.1985	0.0736	0.0736	0.0054	0.960	1.000
Y3		1.000	1.0104	0.0472	0.0474	0.0023	0.950	1.000
E		1.200	1.2607	0.1565	0.1444	0.0279	0.930	1.000
Categorical Latent Variables								
C#1	ON							
C&1#1		0.700	0.6886	0.0529	0.0412	0.0029	0.940	1.000
C&1#2		0.200	0.1929	0.0287	0.0307	0.0009	0.960	1.000

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS  
BASED ON THE ESTIMATED MODEL

Latent Class		Pattern	
1	1	688.63862	0.34432
1	2	311.36138	0.15568
2	1	192.85615	0.09643
2	2	807.14385	0.40357

FINAL CLASS COUNTS AND PROPORTIONS FOR EACH LATENT CLASS VARIABLE  
BASED ON THEIR MOST LIKELY LATENT CLASS PATTERN

Latent Class		Variable	
		Class	
C	1	768	0.38402
	2	1231	0.61598

CLASSIFICATION QUALITY

Entropy 0.835

Average Latent Class Probabilities for Most Likely Latent Class Pattern (Row)  
by Latent Class Pattern (Column)

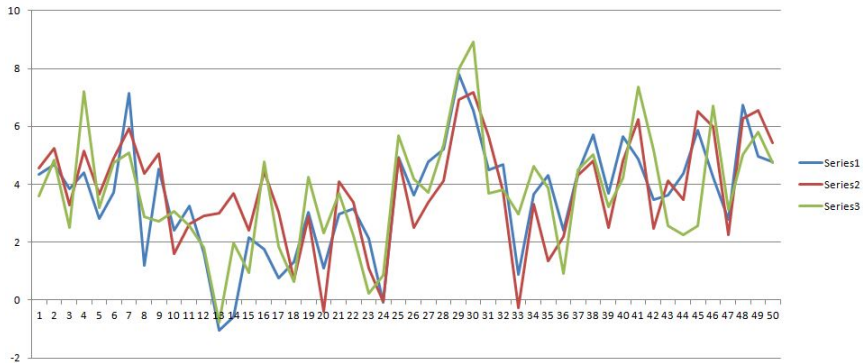
	1	2
1	0.874	0.126
2	0.076	0.924

Small bias in the means. What to do? I left this unanswered on purpose to make the point. Possible potential causes

- Model poorly identified: Not enough parameters differ across class? Simplify model by holding parameter equal to improve identification. Parameters that are not significantly different and make sense can be constrained to be equal. Loadings not significantly different from 1 fix to 1.
- Not enough sample size to get to bias of zero. Increase sample size. Simplify model. Run simulations with bigger sample. Note that Entropy remains the same as sample increases
- Not enough MCMC iterations. Run with more iterations. Look at the traceplot of the offending parameters to evaluate convergence
- Label switching in Mixtures? I have not seen this happen yet. I think it is very rare because of the Markov regime switching/smoothing

- Add informative priors to improve identifiability if the model.
- Maybe entropy is too low?
- In smaller sample size situations the number of regime switching events could be low if the state is stable - not enough to build the model. Recall that we use  $P(C_t = 1|C_{t-1} = 2)$  and  $P(C_t = 2|C_{t-1} = 1)$  to figure out  $C_t$  distribution. If there are very few switch events accurate stable estimates are probably unrealistic.
- Other things I don't know about
- These methods are new and are the cutting edge of methodology. Not using simulation studies in parallel to a real data estimation is irresponsible. Simulation studies not possible in Bugs. They are possible in Mplus because it is much faster.

# MSKF Data



- Class switching are not "clearly" visible. One needs the detailed analysis that the MSKF model provides
- Not very easy to come up with  $E(Y_t|S_t)$  explicitly. Using the simulated data however I can compute that  $E(Y_t|S_t = 1)$  is around 1 and  $E(Y_t|S_t = 2)$  is around 4. This can also be computed by imputing  $S_t$  with the factor score command and computing the MCMC based values.
- Periods 1 to 25: Mostly switching back and forth with a stable phase in regime 1 for periods 10, 11, 12, 13.
- Periods 25 to 50:  $S_t$  has entered a stable regime 2 with only a single drop to regime 1 at period 32. This makes a lot more sense once you know that the two estimated values for  $E(Y_t|S_t)$  are 1 and 4



- Not really two-level models. These are intensive longitudinal models: multiple observations nested within clusters.
- This methodology more suitable for social sciences than the case of  $N=1$
- Dynamic Latent Class Analysis (DLCA)
- Multilevel Markov Switching Autoregressive Models(MMSAR)
- Multilevel Markov Switching Kalman Filter Models (MMSKF)

# Dynamic Latent Class Analysis

- 200 individuals, 4 binary class indicator, 2 class model, each individual has 100 times of observations. The model is traditional LTA with 100 time points

$$P(U_{pit} = 1 | S_{it} = k) = \Phi(\tau_{kp} + \tau_{ip})$$

$$P(S_{it} = 1 | S_{it-1} = 1) = \frac{\text{Exp}(\alpha_{i1})}{1 + \text{Exp}(\alpha_{i1})}$$

$$P(S_{it} = 2 | S_{it-1} = 2) = \frac{\text{Exp}(\alpha_{i2})}{1 + \text{Exp}(\alpha_{i2})}$$

$$\tau_{ip} \sim N(0, \sigma_{ip})$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j)$$

- Subject specific transition matrix

$$\begin{pmatrix} \alpha_{i1} & 1 - \alpha_{i2} \\ 1 - \alpha_{i1} & \alpha_{i2} \end{pmatrix}$$

```
montecarlo:
  names are u1-u4;
  genclasses = c1(2 &1); classes = c1(2 &1);
  nobs = 20000; nrep = 100;
  ncsizes = 1; csizes = 200(100);
  categorical=u1-u4; generate=u1-u4(1);

ANALYSIS:  TYPE = MIXTURE TWOLEVEL RANDOM;
           ESTIMATOR = BAYES; proc=2;

MODEL population:
  %within%
  %OVERALL%
  s1 | C1#1 on C1&1#1; s2 | C1#1 on C1&1#2;
  %BETWEEN%
  %OVERALL%
  s1-s2*0.05; [s1*1 s2*-0.5]; u1-u4*0.1;

MODEL POPULATION-C1:
  %between%
  %c1#1%
  [u1$1-u4$1*1.0];
  %c1#2%
  [u1$1-u4$1*-1.0];
```

# DLCA Results

S2	WITH							
S1		0.000	0.0004	0.0122	0.0127	0.0001	0.900	0.100
Means								
S1		1.000	0.9993	0.0269	0.0295	0.0007	0.970	1.000
S2		-0.500	-0.4958	0.0350	0.0309	0.0012	0.890	1.000
Thresholds								
U1\$1		1.000	0.9984	0.0245	0.0275	0.0006	0.980	1.000
U2\$1		1.000	0.9987	0.0306	0.0280	0.0009	0.940	1.000
U3\$1		1.000	0.9993	0.0321	0.0278	0.0010	0.900	1.000
U4\$1		1.000	0.9940	0.0294	0.0278	0.0009	0.930	1.000
Variances								
U1		0.100	0.1015	0.0151	0.0137	0.0002	0.940	1.000
U2		0.100	0.1035	0.0141	0.0140	0.0002	0.920	1.000
U3		0.100	0.1034	0.0131	0.0139	0.0002	0.960	1.000
U4		0.100	0.1009	0.0127	0.0137	0.0002	0.970	1.000
S1		0.050	0.0564	0.0187	0.0167	0.0004	0.900	1.000
S2		0.050	0.0550	0.0187	0.0180	0.0004	0.920	1.000

# Multilevel Markov Switching Autoregressive Models(MMSAR) Simulation

- 100 individuals with 100 times of observations. One dependent variable  $Y_{it}$
- $S_{it}$  is a two state categorical latent variable.
- We estimate a hidden Markov model for  $S_{it}$  with subject specific transition probabilities
- We estimate a twolevel regression with random intercept and MSAR(1) model for the within part of  $Y_{it}$  (note the mean is on the within level - version 8.1)

$$Y_{it} = Y_{b,i} + Y_{w,it}$$

$$Y_{w,it} = \mu_{S_{it}} + \beta_{S_{it}} Y_{w,it-1} + \varepsilon_{it}$$

$$P(S_{it} = j | S_{it-1} = j) = \frac{\text{Exp}(\alpha_{ij})}{1 + \text{Exp}(\alpha_{ij})}$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j), Y_{b,i} \sim N(0, \sigma)$$

```
montecarlo:
  names are y;
  genclasses = c1(2 &1); classes = c1(2 &1);
  nobs = 10000; ncsizes = 1; csizes = 100(100);
  nrep = 100;

ANALYSIS:  TYPE = MIXTURE TWOLEVEL RANDOM;
           ESTIMATOR = BAYES; proc=2;

MODEL population:
  %within%
  %OVERALL%
  s1 | C1#1 on C1&1#1; s2 | C1#1 on C1&1#2;
  y@0.05; [y@0];
  f by y@1 (&1); f on f&1*0.4;

  %BETWEEN%
  %OVERALL%
  s1-s2*0.05; [s1*1 s2*-0.5]; y*1;

MODEL POPULATION-C1:
  %within%
  %c1#1%
  f on f&1*0.4; f*0.9; [f*1.0];
  %c1#2%
  f on f&1*0.2; f*0.7; [f*-1.0];
```

# MMSAR Results

```
Within Level
Latent Class 1
F          ON
  F&l      0.400    0.3919    0.0129    0.0146    0.0002 0.939 1.000
Intercepts
  F        1.000    0.9902    0.0785    0.0572    0.0062 0.768 1.000
Residual Variances
  F        0.900    0.9104    0.0360    0.0362    0.0014 0.939 1.000
Latent Class 2
F          ON
  F&l      0.200    0.1955    0.0139    0.0142    0.0002 0.939 1.000
Intercepts
  F       -1.000   -1.0234    0.0993    0.0670    0.0103 0.737 1.000
Residual Variances
  F        0.700    0.7068    0.0369    0.0337    0.0014 0.939 1.000
Between Level
Means
  S1       1.000    1.0160    0.0696    0.0671    0.0050 0.919 1.000
  S2      -0.500   -0.5136    0.0807    0.0883    0.0066 0.939 1.000
Variances
  Y        1.000    1.0314    0.1575    0.1594    0.0255 0.939 1.000
  S1       0.050    0.0613    0.0292    0.0341    0.0010 0.949 1.000
  S2       0.050    0.0724    0.0411    0.0449    0.0022 0.929 1.000
```

# Multilevel Markov Switching Kalman Filter Models (MMSKF)

- 100 individuals with 100 times of observations, 4 continuous factor indicators  $Y_{pit}$  measuring a factor  $\eta_{it}$ .
- $S_{it}$  is a two state categorical latent variable.
- We estimate a hidden Markov model for  $S_t$  with subject specific transition probabilities
- MSAR(1) model for the factor
- For  $p=1, \dots, 4$

$$Y_{pit} = v_{pi} + \lambda_p \eta_{it} + \varepsilon_{pit}$$

$$\eta_{it} = \mu_{S_{it}} + \beta_{S_{it}} \eta_{it-1} + \xi_{it}$$

$$P(S_{it} = j | S_{it-1} = j) = \frac{\text{Exp}(\alpha_{ij})}{1 + \text{Exp}(\alpha_{ij})}$$

$$\alpha_{ij} \sim N(\alpha_j, \sigma_j)$$

$$v_{pi} \sim N(v_p, \sigma_p)$$

- For identification purposes  $\mu_1 = 0$  and  $\lambda_1 = 1$



```
montecarlo: names are u1-u4;
  genclasses = c1(2 &1); classes = c1(2 &1);
  nobobs = 10000; ncsizes = 1; csizes = 100(100); nrep = 100;

ANALYSIS: TYPE = MIXTURE TWOLEVEL RANDOM; ESTIMATOR = BAYES; proc=2;

MODEL population:
  %within%
  %OVERALL%
  s1 | C1#1 on C1&1#1; s2 | C1#1 on C1&1#2;
  f by u1@1 u2*1.2 u3-u4*0.8 (&1); f on f&1*0.4; f*1; u1-u4*1;
  %BETWEEN%
  %OVERALL%
  s1-s2*0.05; [s1*1 s2*-0.5]; u1-u4*1;

MODEL POPULATION-C1:
  %within%
  %c1#1%
  f on f&1*0.4; f*1; [f@0]; [u1-u4*0] (m1-m4);
  %c1#2%
  f on f&1*0.2; f*0.8; [f*2.5]; [u1-u4*0] (m1-m4);
```

# MMSKF Results, within level, class 1

		Population	ESTIMATES		S. E.	M. S. E.	95% Cover	% Sig
			Average	Std. Dev.	Average			Coeff
Within Level								
Latent Class 1								
F	BY							
U1		1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
U2		1.200	1.2022	0.0118	0.0109	0.0001	0.940	1.000
U3		0.800	0.8016	0.0085	0.0087	0.0001	0.950	1.000
U4		0.800	0.8010	0.0094	0.0086	0.0001	0.950	1.000
F	ON							
F&1		0.400	0.3935	0.0170	0.0172	0.0003	0.930	1.000
Intercepts								
U1		0.000	0.0197	0.1166	0.1193	0.0138	0.950	0.050
U2		0.000	-0.0015	0.1244	0.1248	0.0153	0.950	0.050
U3		0.000	0.0049	0.1166	0.1139	0.0135	0.960	0.040
U4		0.000	0.0023	0.1186	0.1127	0.0139	0.930	0.070
F		0.000	0.0000	0.0000	0.0000	0.0000	1.000	0.000
Residual Variances								
U1		1.000	1.0010	0.0197	0.0194	0.0004	0.950	1.000
U2		1.000	0.9968	0.0221	0.0227	0.0005	0.970	1.000
U3		1.000	1.0010	0.0163	0.0170	0.0003	0.950	1.000
U4		1.000	1.0009	0.0170	0.0171	0.0003	0.960	1.000
F		1.000	1.0125	0.0534	0.0525	0.0030	0.940	1.000

# MMSKF Results, within level, class 2

Latent Class 2								
F	BY							
U1		1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
U2		1.200	1.2022	0.0118	0.0109	0.0001	0.940	1.000
U3		0.800	0.8016	0.0085	0.0087	0.0001	0.950	1.000
U4		0.800	0.8010	0.0094	0.0086	0.0001	0.950	1.000
F	ON							
F&1		0.200	0.1962	0.0186	0.0176	0.0004	0.940	1.000
Intercepts								
U1		0.000	0.0197	0.1166	0.1193	0.0138	0.950	0.050
U2		0.000	-0.0015	0.1244	0.1248	0.0153	0.950	0.050
U3		0.000	0.0049	0.1166	0.1139	0.0135	0.960	0.040
U4		0.000	0.0023	0.1186	0.1127	0.0139	0.930	0.070
F		2.500	2.4932	0.0531	0.0494	0.0028	0.940	1.000
Residual Variances								
U1		1.000	1.0010	0.0197	0.0194	0.0004	0.950	1.000
U2		1.000	0.9968	0.0221	0.0227	0.0005	0.970	1.000
U3		1.000	1.0010	0.0163	0.0170	0.0003	0.950	1.000
U4		1.000	1.0009	0.0170	0.0171	0.0003	0.960	1.000
F		0.800	0.8158	0.0572	0.0521	0.0035	0.920	1.000

# MMSKF Results, between level

## Between Level

### Means

S1	1.000	1.0134	0.0856	0.0782	0.0074	0.920	1.000
S2	-0.500	-0.5193	0.1197	0.1053	0.0146	0.870	0.980

### Variances

U1	1.000	1.0456	0.1736	0.1644	0.0319	0.930	1.000
U2	1.000	1.0168	0.1524	0.1638	0.0233	0.970	1.000
U3	1.000	1.0620	0.1539	0.1642	0.0273	0.920	1.000
U4	1.000	1.0300	0.1578	0.1592	0.0255	0.970	1.000
S1	0.050	0.0641	0.0349	0.0363	0.0014	0.930	1.000
S2	0.050	0.0826	0.0437	0.0518	0.0030	0.920	1.000

- Determine the number of classes: ignore time series and use standard methods
- Starting values: maybe coming soon, not as big issue for Bayes as it is for ML due to MCMC naturally goes through many starting values
- Comparing models: maybe coming soon DIC, model test, new PPP methods, other new methods
- Multiple solutions: We get a lot of multiple solutions with ML. Does that happen with Bayes too? Using different starting values? This is an issue even without time-series.
- Label switching - hopefully not much of a problem due to Markov smoothing