Estimating new structural equation models with the Bayesian methods
(future ideas for Mplus and time series modeling)

Tihomir Asparouhov
Mplus

August 30, 2012

Overview

- Overview of time series models and how to run some time series models in Mplus
- Overview of time series models and SEM, dynamic factor analysis (DFA) models and Kalman filter
- How to incorporate DFA and Kalman filter in Mplus with Bayes estimation

Overview of time series models

- Time series models: analysis of data measured at successive time instants
- Wide and solid applications: econometrics, signal processing, and mathematical finance
- Intensive longitudinal data (ILD) in the social sciences: more longitudinal data are collected that makes very frequent observations using new tools for data collection such as palm pilots, smart phones etc. Walls & Schafer (2006)
- Jahng S., Wood, P. K.,& Trull, T. J., (2008) Analysis of Affective Instability in Ecological Momentary Assessment: Indices Using Successive Difference and Group Comparison via Multilevel Modeling. Psychological Methods, 13, 354-375
- Measurement instrument for a "mood" factor collected several times a day for several months
- Ecological Momentary Assessment (EMA): involves repeated sampling of subjects current behaviors and experiences in real time, in subjects natural environments

Overview of time series models

- New book: Bolger & Laurenceau (2012), Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research New York: Guilford Press
- Uses Mplus prominently and includes examples with Mplus inputs

Overview of time series models

- Time series models are models for the disturbances/residuals in a model as a function of time
- It is easy to combine with SEM for any longitudinal data. Two separate models: a **Structural model** and **Disturbances model = Time series model**
- Any residual variable can be modeled further (beyond the SEM part) with a time series model
- Modeling frameworks that combine SEM and Time-series: Dynamic Factor Analysis and Kalman Filter model

Overview of time series models

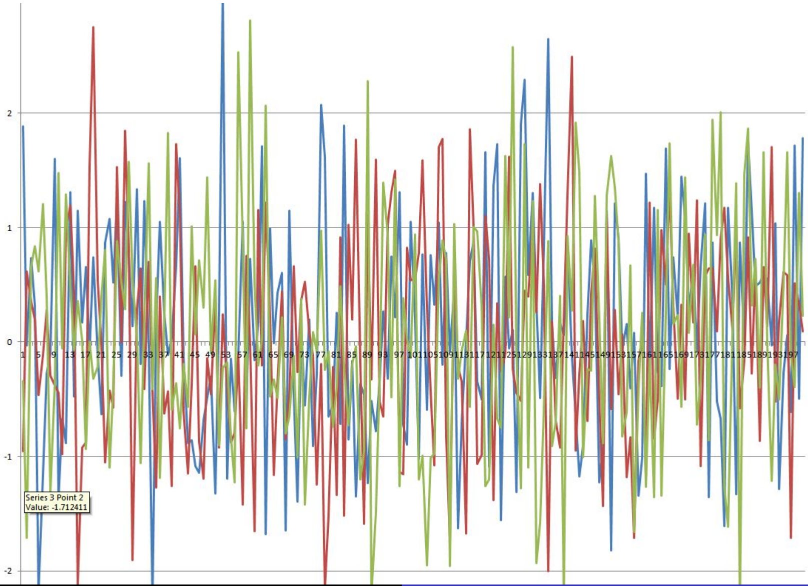
Table : Google search results in millions

Term	Pages
Time series	22.8
Factor analysis	4.3
Principal component analysis	4.3
Kalman filter	3.7
Mplus	1.7
Multiple imputation	0.25
Dynamic factor analysis	0.09

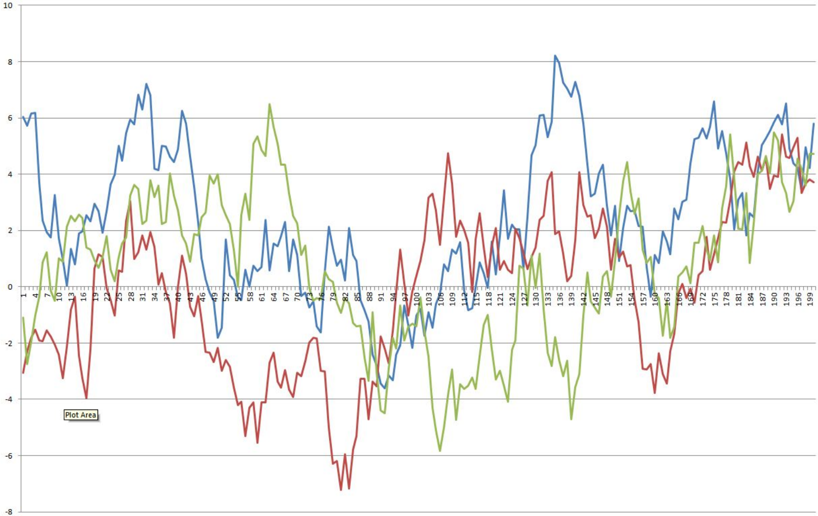
Overview of time series models

- We generate data on 100 individuals with 200 time points
- Using a Kalman filter model with 5 indicators 1 factor
- Two plots of within level factor, using AR(1)
- Using autoregressive coefficient of 0.95 and 0
- Both plots represent zero mean within factor value with the same variance (disturbance process) (the between factor is not included at time *t* or any growth/trend values)
- Four colors represent 4 individuals/clusters

Residual plot with autocorrelation of 0



Residual plot with autocorrelation of 0.95



Why model the disturbances?

- Without time series modeling we are assuming picture 1 where the reality maybe picture 2
- Without time series modeling we fail to understand the process behind the data
- Ignoring the correlations between the residuals will underestimate the SE
- The predictive power of a model without the time series model will be worse

Disturbance models - AR(1)

- AR(1) - autoregressive model with lag 1, with autocorrelation ρ

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t$$

- For $t > 1$, $Var(\xi_t) = \sigma^2$
- For the first term which is not regressed on anything, $Var(\xi_1) = \sigma^2/(1-\rho^2)$
- Stationary process, i.e., the variance of the disturbance is constant across time $Var(\varepsilon_t) = \sigma^2/(1-\rho^2)$

Disturbance models - AR(1)

- implies a correlation matrix for ε_t, $t = 1, ..., T$

$$\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & ... & \rho^T \\ \rho & 1 & \rho & \rho^2 & ... & \rho^{T-1} \\ ... & ... & ... & ... & ... & ... \\ \rho^T & \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & ... & 1 \end{pmatrix}$$

- $Corr(\varepsilon_t, \varepsilon_s) = \rho^{t-s}$

Disturbance models - AR(1). How to run in Mplus - Method 1 as in UG example 6.17

```
MODEL:                y1-y4 (resvar);
      y1-y3 PWITH y2-y4 (p1);
      y1-y2 PWITH y3-y4 (p2);
      y1 WITH y4 (p3);
MODEL CONSTRAINT:
      NEW (corr);
      p1 = resvar*corr;
      p2 = resvar*corr**2;
      p3 = resvar*corr**3;
```

Disturbance models - AR(1). How to run in Mplus - Method 2

```
MODEL:
y2-y4 pon y1-y3 (corr);
y2-y4 (v); y1 (v1);
MODEL CONSTRAINT:
v1=v/(1-corr*corr);
```

Disturbance models - AR(1). Comparison: Method 1 v.s. Method 2

```
Loglikelihood

          H0 Value                          -6682.537

MODEL RESULTS
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 New/Additional Parameters
    CORR               0.718      0.010     68.816      0.000
```

```
Loglikelihood

          H0 Value                          -6682.537

MODEL RESULTS
                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 Y2         ON
    Y1                 0.718      0.010     68.816      0.000
```

Disturbance models - AR(1). Simulation study with T=100

```
MONTECARLO:
      NAMES = y1-y100;
      NOBS = 100;
      NREPS = 100;

MODEL POPULATION:
y2-y100 pon y1-y99*0.7 (corr);
y2-y100*1 (v); y1*1.96 (v1);
[y1-y100*0] (m);

MODEL:
y2-y100 pon y1-y99*0.7 (corr);
y2-y100*1 (v); y1*1.96 (v1);
[y1-y100*0] (m);

MODEL CONSTRAINT:
v1=v/(1-corr*corr);
```

Disturbance models - AR(1). Simulation study results, T=100

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
Y2 ON							
Y1	0.700	0.7000	0.0072	0.0071	0.0001	0.940	1.000
Means							
Y1	0.000	0.0012	0.0098	0.0100	0.0001	0.970	0.030
Variances							
Y1	1.960	1.9632	0.0441	0.0470	0.0019	0.970	1.000
Y2	1.000	1.0009	0.0131	0.0142	0.0002	0.980	1.000
Y3	1.000	1.0009	0.0131	0.0142	0.0002	0.980	1.000

Disturbance models - AR(1). Limitations of multivariate setup

- It works with T=200, with T=300 or more not computable due to memory problems.
- If the model is multivariate then maximum possible T will have to be divided by P (the number of variables)
- If T is large (T=10000 in the next example) the model can be run as a univariate model where the data is organized in a long format with the lag 1 variables set as covariate $Y1$. Ignore the initial equation.
- For large T ignoring the initial equation has negligible effect on the estimation.
- It is important to set $Y1$ as covariate to preserve log-likelihood value

Disturbance models - AR(1). Univariate setup Method 3: duplication

Data setup

0.20510475	-0.378137
-0.590985813	0.20510475
-1.027777453	-0.590985813
-2.863705363	-1.027777453
1.674698659	-2.863705363
0.788097665	1.674698659
0.766737416	0.788097665
2.031961354	0.766737416
2.394893339	2.031961354
0.050893335	2.394893339
-0.095748666	0.050893335
1.094087833	-0.095748666

Disturbance models - AR(1). Univariate setup Method 3: duplication

```
variable:
      names = y y1;

data: file = dbl.dat;

model:
y on y1*0.25;
```

Disturbance models - AR(1). Univariate setup Method 3: results

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Y ON					
	Y1	0.269	0.010	27.873	0.000
Intercepts					
	Y	-0.002	0.010	-0.189	0.850
Residual Variances					
	Y	1.026	0.015	70.700	0.000

Disturbance models - AR(1)

- The AR(1) is a model for the disturbances. How do you setup the model in the presence of Y predictors such as latent variables from a growth model or observed covariates

```
MODEL:      i s | y1@0 y2@1 y3@2 y4@3;
      y1-y4 (resvar);
      y1-y3 PWITH y2-y4 (p1);
      y1-y2 PWITH y3-y4 (p2);
      y1 WITH y4 (p3);
MODEL CONSTRAINT:
      NEW (corr);
      p1 = resvar*corr;
      p2 = resvar*corr**2;
      p3 = resvar*corr**3;
```

Disturbance models - AR(1) for growth model Method 2

```
MODEL:     i s | y1@0 y2@1 y3@2 y4@3;
      y1-y4@0;
      f1 by y1@1;
      f2 by y2@1;
      f3 by y3@1;
      f4 by y4@1;
      f2-f4 pon f1-f3 (corr);
      f2-f4 (v); f1 (v1);
MODEL CONSTRAINT:
v1=v/(1-corr*corr);
```

Disturbance models - more models

- MA(1) - moving average model with lag 1

$$\varepsilon_t = \xi_t + \theta\xi_{t-1}$$

$Cov(\varepsilon_t, \varepsilon_s)$=0 if $t - s > 1$

- ARMA(1,1) - autoregressive model with lag 1 and moving average model with lag 1

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t + \theta\xi_{t-1}$$

- ARMA(p,q) - autoregressive model with lag 1 and moving average model with lag 1

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + ... + \rho_p\varepsilon_{t-p} + \xi_t + \theta_1\xi_{t-1} + ... + \theta_q\xi_{t-q}$$

- ARMA(p,q) models have been very successful disturbance models in practice

Disturbance models - ARMA(1,1) setup in Mplus, T=10

```
MODEL:
y2-y10 pon y1-y9*0.7 (corr);
[y1-y10*0] (m);
y2-y10@0;
y1*1.04 (v1);

f0-f10 with f0-f10@0;
f0-f10*1(v);
f0 by       y1*0.3 (theta);
f1 by y1@1 y2*0.3 (theta);
f2 by y2@1 y3*0.3 (theta);
......
f9 by y9@1 y10*0.3 (theta);
f10 by y10@1;

MODEL CONSTRAINT:
v1=corr*corr*v*(1+theta*theta)/(1-corr*corr);
```

DFA and Kalman filter

- Kalman filter$\approx$ DFA $\approx$ ARMA disturbance model for the factor over time.
- Molenaar (1985). A dynamic factor model for the analysis of multivariate time series. Psychometrika
- Zhang, Hamaker and Nesselroade (2008) Comparisons of Four Methods for Estimating a Dynamic Factor Model. Structural Equation Modeling
- Zhang and Nesselroade (2007) Bayesian Estimation of Categorical Dynamic Factor Models
- Justiniano (2004). Estimation and model selection in dynamic factor analysis. PhD dissertation

Two special DFA models

- Direct autoregressive factor score (DAFS)

$$Y_t = \Lambda f_t + \varepsilon_t$$

$$f_t = \rho f_{t-1} + \xi_t$$

- White noise factor score (WNFS)

$$Y_t = \Lambda f_t + \Lambda_1 f_{t-1} + \varepsilon_t$$

DAFS model estimation in Mplus following Molenaar (1985): duplication

- Suppose that we have 3 indicator 1 factor DAFS model
- Organize the data as follows: duplication
  Y1 Y2 Y3 YLag11 YLag12 YLag13
  1 2 3 4 5 6
  7 8 9 1 2 3
  10 12 12 7 8 9
  ....

DAFS model estimation in Mplus following Molenaar (1985): duplication

- write the model as

```
f by Y1-Y3 (l1-l3); f@1;
flag1 by YLag11-YLag13 (l1-l3); flag1@1;
[Y1-Y3](m1-m3);
[YLag11-YLag13](m1-m3);
Y1-Y3(v1-v3);
YLag11-YLag13(v1-v3);
f on flag1 (rho);
```

- This is not the original suggestion by Molenaar but the above setup yields the same likelihood

DAFS model estimation in Mplus following Molenaar (1985): duplication

- Problem 1: observations are not independent
- Problem 2: tests of fit are meaningless
- Problem 3: in AR(1) regression equation only observed data from lag 1 have effect
- Problem 4: Standard error would have to be adjusted
- Problem 5: Not true ML
- Problem 6: simulations show that it works well in some cases but no guarantee that will work well in all cases

Bayes estimation for DFA models

- Consider the model

$$Y_{ti} = Y_{B,i} + \Lambda f_{ti} + \varepsilon_{ti}$$

$$f_{ti} = f_{B,i} + f_{W,ti}$$

$$f_{W,ti} = \rho f_{W,t-1,i} + \xi_{ti}$$

  for i=1,...T and i=1,...,N
- This model has a two-level factor structure with individually specific factor mean and AR(1) structure for the disturbances
- This model is more general than DAFS and WNFS because it adds two-level structure
- One way to think about this model is that the two-level part explains the mean of f_{ti} within a cluster while the AR(1) explains the covariance within a cluster and abandons the assumption that observations are conditionally independent given the between level random effects

Bayes estimation for DFA models

If we ignore the AR(1) part of the model, Mplus MCMC will take the following steps to generate parameters and latent variable

- Step 1.

  $$[f_{B,i}, Y_{B,i} | Y_{ti}, parameters]$$

  multivariate normal

- New Step 1

  $$[f_{B,i}, Y_{B,i} | f_{W,ti}, Y_{ti}, parameters]$$

  conditioning on $f_{W,ti}$ makes this a feasible step otherwise the $Y_{ti}|$ between level random effects are not independent, loss of efficiency

Bayes estimation for DFA models

- Step 2. $[f_{w,it}|f_{B,i}, Y_{ti}, parameters]$ simply subtract the between parts and conduct independent factor analysis update for each $f_{w,it}$
- New Step 2. $[f_{w,it}|f_{B,i}, Y_{ti}, parameters]$ has to be broken into T steps

$$[f_{w,i1}|f_{w,i,-1}f_{B,i}, Y_{ti}, parameters]$$

$$[f_{w,i2}|f_{w,i,-2}f_{B,i}, Y_{ti}, parameters]$$

$$............$$

$$[f_{w,iT}|f_{w,i,-T}f_{B,i}, Y_{ti}, parameters]$$

where $f_{w,i,-t}$ means all within level factors but $f_{w,i,t}$, i.e., $f_{w,i,1}$,... ,$f_{w,i,t-1}$,$f_{w,i,t+1}$,...,$f_{w,i,T}$

Bayes estimation for DFA models

- New Step 2.

$$[f_{w,it}|f_{w,i,-t}f_{B,i}, Y_{ti}, parameters]$$

is based on the following equations

$$Y_{W,ti} = \Lambda f_{W,ti} + \varepsilon_{ti}$$

$$f_{W,t+1,i} = \rho f_{W,ti} + \xi_{t+1,i}$$

$$f_{W,ti} = \rho f_{W,t-1,i} + \xi_{ti}$$

- The factor that we need to update has one more indicator $f_{W,t+1,i}$ and a predictor $f_{W,t-1,i}$
- For t=1 and t=T the model looses one of the two new equations. It is important to properly account for the ending equations probably for $T < 200$.
- For t=1, $f_{W,ti} = \sigma^2/(1-\rho^2)$. It is important to generate $f_{W,1i}$ with this bigger variance otherwise the Baysian generation of $f_{W,ti}$ will not look stationary and would have an increasing variance.

Bayes estimation for DFA models

- The estimation of slopes, intercepts, and loadings, given the latent variables is identical, including ρ, however the estimation of ρ is based on one fewer equations than all other parameters: $T-1$ equations, which actually causes change in the use of sufficient statistics

$$f_{W,ti} = \rho f_{W,t-1,i} + \xi_{ti}$$

- The estimation of the variance parameters is the same with one exception: $Var(\xi_{ti})$. For $t > 1$, $Var(\xi_{ti}) = \sigma^2$ but for $t = 1$ $Var(\xi_{ti}) = \sigma^2/(1-\rho^2)$. To resolve this problem in updating σ^2 we use $\sqrt{(1-\rho^2)}\xi_{1i}$ instead of ξ_{1i}.
- The convergence will be fast as long as the factor is measured well by the indicators, that would make all variables nearly observed.

Bayes estimation for DFA models

```
%within%
f by y1-y5;
f on f%1; ! AR(1) disturbance model
%between%
fb by y1-y5;
```

Bayes estimation for DFA models: Conclusions

- Developing disturbance models in Mplus will combine SEM framework with the time series framework and deliver many new modeling possibilities
- Consider time intense data Y_{it} analyzed as cross-classified, crossed by individual and time. The variance decomposition is

$$Var(Y_{it}) = S_i + S_t + S_{it}$$

  and now we would be able to add autoregressive structure on the within level as well.