An Efficient State Space Approach to Estimate Univariate and Multivariate Multilevel Regression Models

Fei Gu
Kristopher J. Preacher
Wei Wu
05/21/2013

Overview

- Introduction: estimate MLM as SEM (Bauer, 2003; Curran, 2003; Mehta & Neale, 2005)
- Multiple-subject state space model (SSM)
- Examples of the SSM approach to

1. Univariate MLM: random intercept and random slope model
2. Univariate MLM: random intercept longitudinal model
3. Multivariate MLM: random intercepts model

- Summary, Discussions, and Conclusions

Introduction

- The SEM approach to MLMs: transforming data set structure (long format to wide format; e.g.,Bauer, 2003; Curran, 2003)
- Equivalence between the SEM and the SSM can give rise to the state space approach conforming the SEM approach
- Limitations: computationally inefficient; “data management nightmare” (Curran, 2003, p. 565)

Introduction

- The state space approach in the time series literature: Icaza & Jones, 1999; Jones, 1993
- Mentioned in educational and psychological literature briefly by Ho, Shumway, & Ombao, 2006, pp. 153-154

State space formulation

$$y_i = \tau + \Lambda \eta_i + \Gamma x_i + \varepsilon_i, \qquad \varepsilon_i \sim MVN(0, \Theta),$$

$$\eta_i = \alpha + \mathrm{B} \eta_{i-1} + \Upsilon x_i + \zeta_i, \qquad \zeta_i \sim MVN(0, \Psi).$$

- y_i: p-variate data vector
- η_i: q-variate latent state vector
- x_i: g-variate vector containing exogenous variables
- ε_i: $p \times 1$ vector containing measurement errors
- ζ_i: $q \times 1$ vector containing latent state noises
- subscript, i, (i = 1, 2, ..., T) denotes the time point

Parameter estimation

- Kalman filter (Kalman, 1960)

$$
\begin{aligned}
\eta_{i|i-1} &= \alpha + \mathrm{B}\,\eta_{i-1|i-1} + \Upsilon x_i \\
\mathrm{P}_{i|i-1} &= \mathrm{B}\mathrm{P}_{i-1|i-1}\mathrm{B}' + \Psi \\
e_i &= y_i - y_{i|i-1} = y_i - (\tau + \Lambda\,\eta_{i|i-1} + \Gamma x_i) \\
\mathrm{D}_i &= \Lambda \mathrm{P}_{i|i-1}\Lambda' + \Theta \\
\mathrm{K}_i &= \mathrm{P}_{i|i-1}\Lambda' \mathrm{D}_i^{-1} \\
\eta_{i|i} &= \eta_{i|i-1} + \mathrm{K}_i e_i = \eta_{i|i-1} + \mathrm{P}_{i|i-1}\Lambda' \mathrm{D}_i^{-1} e_i \\
\mathrm{P}_{i|i} &= \mathrm{P}_{i|i-1} - \mathrm{K}_i \mathrm{D}_i \mathrm{K}_i' = \mathrm{P}_{i|i-1} - \mathrm{P}_{i|i-1}\Lambda' \mathrm{D}_i^{-1}\Lambda \mathrm{P}_{i|i-1}.
\end{aligned}
$$

- Prediction Error Decomposition (PED; Schweppe, 1965) function:

$$
\mathrm{PED} = \frac{1}{2}\sum_{i=1}^{\mathrm{T}}\left[-p\log(2\pi) - \log\left|\mathrm{D}_i\right| - e_i'\mathrm{D}_i^{-1}e_i\right].
$$

Multiple-subject SSM

- A second subscript, *j*, (*j* = 1, 2, …, J) indexing different units is added to both the measurement and transition equations:

$$y_{ij} = \tau_{\cdot j} + \Lambda_{\cdot j}\eta_{ij} + \Gamma_{\cdot j}x_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim MVN(0, \Theta_{\cdot j}),$$

$$\eta_{ij} = \alpha_{\cdot j} + \mathrm{B}_{\cdot j}\eta_{i-1,j} + \Upsilon_{\cdot j}x_{ij} + \zeta_{ij}, \qquad \zeta_{ij} \sim MVN(0, \Psi_{\cdot j}).$$

- Total PED function:

$$\text{total PED} = \frac{1}{2}\sum_{j=1}^{\mathrm{J}}\sum_{i=1}^{\mathrm{T}}\left[-p_j \log(2\pi) - \log\left|\mathrm{D}_{ij}\right| - e'_{ij}\mathrm{D}_{ij}^{-1}e_{ij}\right].$$

Example 1: univariate MLM (random intercept and random slope model)

- Level-1: $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon}^{2}\right)$
- Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + u_{0j}$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_{1j} + u_{1j},$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

- Combined equation:

$$y_{ij} = \left[\gamma_{00} + \gamma_{01} w_{1j} + \gamma_{10} x_{ij} + \gamma_{11} w_{1j} x_{ij} \right] + \left[u_{0j} + u_{1j} x_{ij} \right] + \varepsilon_{ij}.$$

Matrix form of the MLM

- Write the combined equation in Laird & Ware (1982) form:

$$\begin{bmatrix} y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & w_{1j} & x_{ij} & w_{1j}x_{ij} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} 1 & x_{ij} \end{bmatrix} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij} \end{bmatrix}.$$

- Rearrange the terms to match the measurement equation of the SSM:

$$\begin{bmatrix} y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} + \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix} \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j}x_{ij} \end{bmatrix} + \begin{bmatrix} \varepsilon_{ij} \end{bmatrix}.$$

Transition equation and the state space form for the MLM

- Within a cluster/group, we have the transition equation:
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}.$$
- The state space form is obtained by define

$$y_{ij} = \begin{bmatrix} y_{ij} \end{bmatrix}, \quad \eta_{ij} = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \quad x_{ij} = \begin{bmatrix} 1 \\ w_{1j} \\ x_{ij} \\ w_{1j} x_{ij} \end{bmatrix}, \quad \varepsilon_{ij} = \begin{bmatrix} \varepsilon_{ij} \end{bmatrix}, \quad \zeta_{ij} = [0],$$

$$\tau_{\cdot j} = [0], \quad \Lambda_{\cdot j} = \begin{bmatrix} 1 & x_{ij} \end{bmatrix}, \quad \Gamma_{\cdot j} = \begin{bmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \end{bmatrix}, \quad \Theta_{\cdot j} = \begin{bmatrix} \sigma_{\varepsilon}^2 \end{bmatrix},$$

$$\alpha_{\cdot j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathrm{B}_{\cdot j} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Upsilon_{\cdot j} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \Psi_{\cdot j} = \begin{bmatrix} 0 & \\ 0 & 0 \end{bmatrix}.$$

Initializing the KF

- To capture the random effects from the level-2 units:

$$\eta_{0|0,j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathrm{P}_{0|0,j} = \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}$$

- Ready to estimate the MLM using a simulated data set:

50 level-2 units; unbalanced design with 12 – 30 level-1 units in each cluster (Poisson(20))

Example 1 results

Parameter	True value	MLwiN		PROC MIXED		PROC IML	
		Point	SE	Point	SE	Point	SE
γ_{00}	10	10.144	.704	10.146	.704	10.146	.704
γ_{01}	8	7.679	.866	7.679	.866	7.679	.866
γ_{10}	5	5.224	.611	5.226	.611	5.226	.611
γ_{11}	4	3.675	.752	3.674	.752	3.674	.752
σ_{ε}^2	1	1.063	.049	1.063	.049	1.063	.049
τ_{00}	10	8.182	1.689	8.187	1.678	8.188	1.678
τ_{10}	4	2.796	1.106	2.800	1.104	2.800	1.104
τ_{11}	8	6.341	1.266	6.339	1.269	6.339	1.270
-2*loglik		3636.527		3636.530		3636.530	

Computational inefficiency

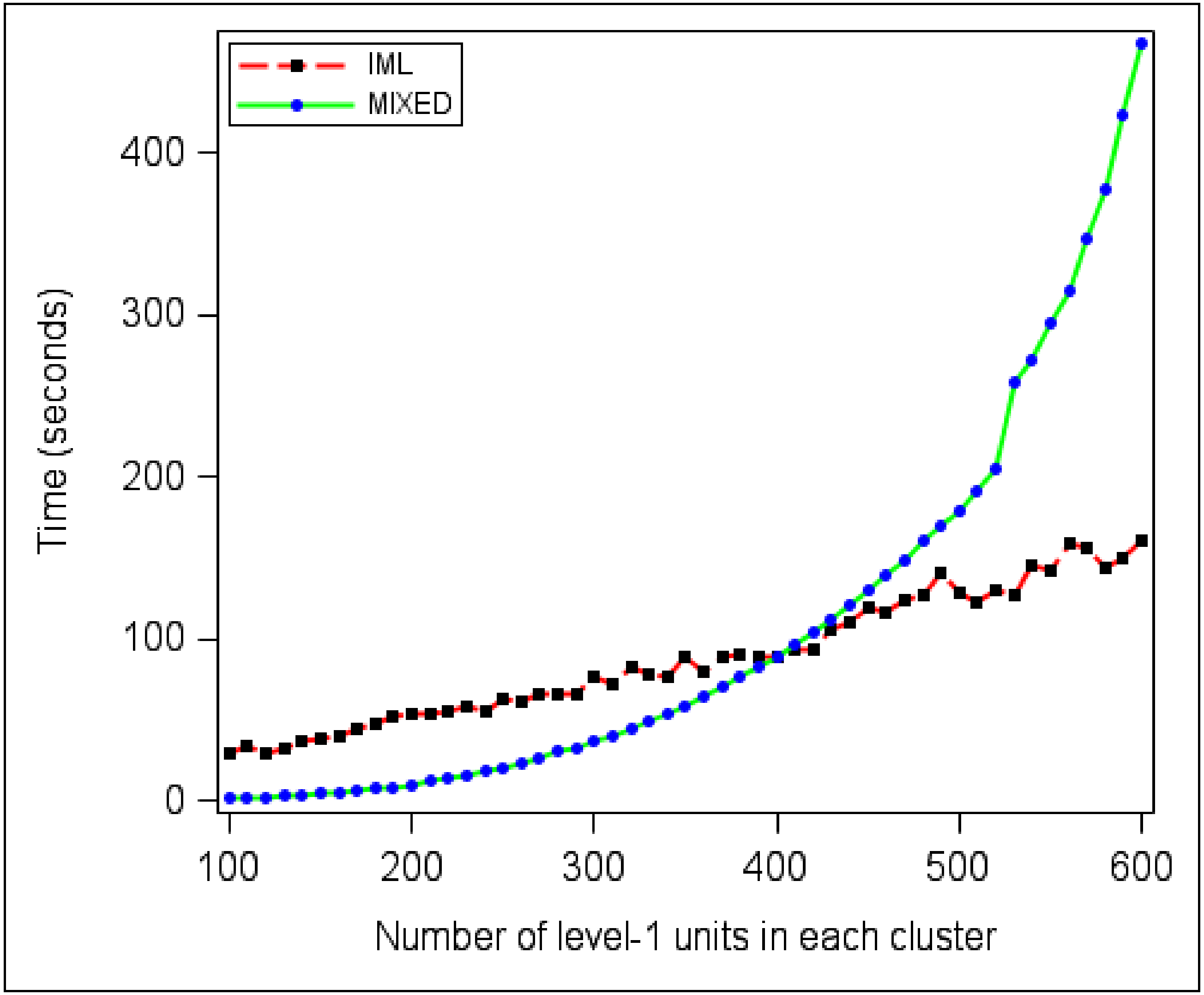
- Estimators specifically designed for MLMs involve matrix operations at the cluster level.
- The dimension of the covariance matrix to be inverted is determined by the number of level-1 units within each level-2 unit.

Computational efficiency

- In the KF, D_i (or D_{ij}) is the only matrix that needs to be inverted in the (multiple-subject) Kalman recursion. The dimension of D_i (or D_{ij}) corresponds to that of the data vector, p.
- In this example, D_i (or D_{ij}) reduces to a scalar, rendering the computations very efficient.

Empirical comparison

- 50 clusters, balanced designs



Example 2: univariate MLM (random intercept longitudinal model)

- The longitudinal model (Icaza & Jones, 1999):

$$y_{ij} = \mu + \gamma_j + \varepsilon_{ij} + \nu_{ij}, \quad \text{where } \gamma_j \sim N(0, \tau), \quad \nu_{ij} \sim N(0, \sigma_\nu^2),$$

$$\varepsilon_{ij} = \phi \varepsilon_{i-1,j} + u_{ij}, \quad \text{where } u_{ij} \sim N(0, \sigma_u^2),$$

- The state space form specification:

$$y_{ij} = \left[y_{ij} \right], \quad \eta_{ij} = \begin{bmatrix} \varepsilon_{ij} \\ \gamma_j \end{bmatrix}, \quad x_{ij} = [1], \quad \varepsilon_{ij} = \left[\nu_{ij} \right], \quad \zeta_{ij} = [0],$$

$$\tau_{\cdot j} = [0], \quad \Lambda_{\cdot j} = [1 \quad 1], \quad \Gamma_{\cdot j} = [\mu], \quad \Theta_{\cdot j} = \left[\sigma_\nu^2 \right],$$

$$\alpha_{\cdot j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathrm{B}_{\cdot j} = \begin{bmatrix} \phi & 0 \\ 0 & 1 \end{bmatrix}, \quad \Upsilon_{\cdot j} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \Psi_{\cdot j} = \begin{bmatrix} \sigma_u^2 & \\ 0 & 0 \end{bmatrix},$$

Initializing the KF

- Initialized with

$$\eta_{0|0,j} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathrm{P}_{0|0,j} = \begin{bmatrix} \dfrac{\sigma_u^2}{1-\phi^2} & \\ 0 & \tau \end{bmatrix}$$

for all *j* to capture the random effect and the innovation variance of the AR(1) process.

- Note

$$\mathrm{Var}(\varepsilon_{ij}) = \phi^2 \mathrm{Var}(\varepsilon_{i-1,j}) + \mathrm{Var}(u_{ij}) \quad \Rightarrow \quad \mathrm{Var}(\varepsilon_{ij}) = \mathrm{Var}(\varepsilon_{i-1,j}) = \frac{\sigma_u^2}{1-\phi^2}.$$

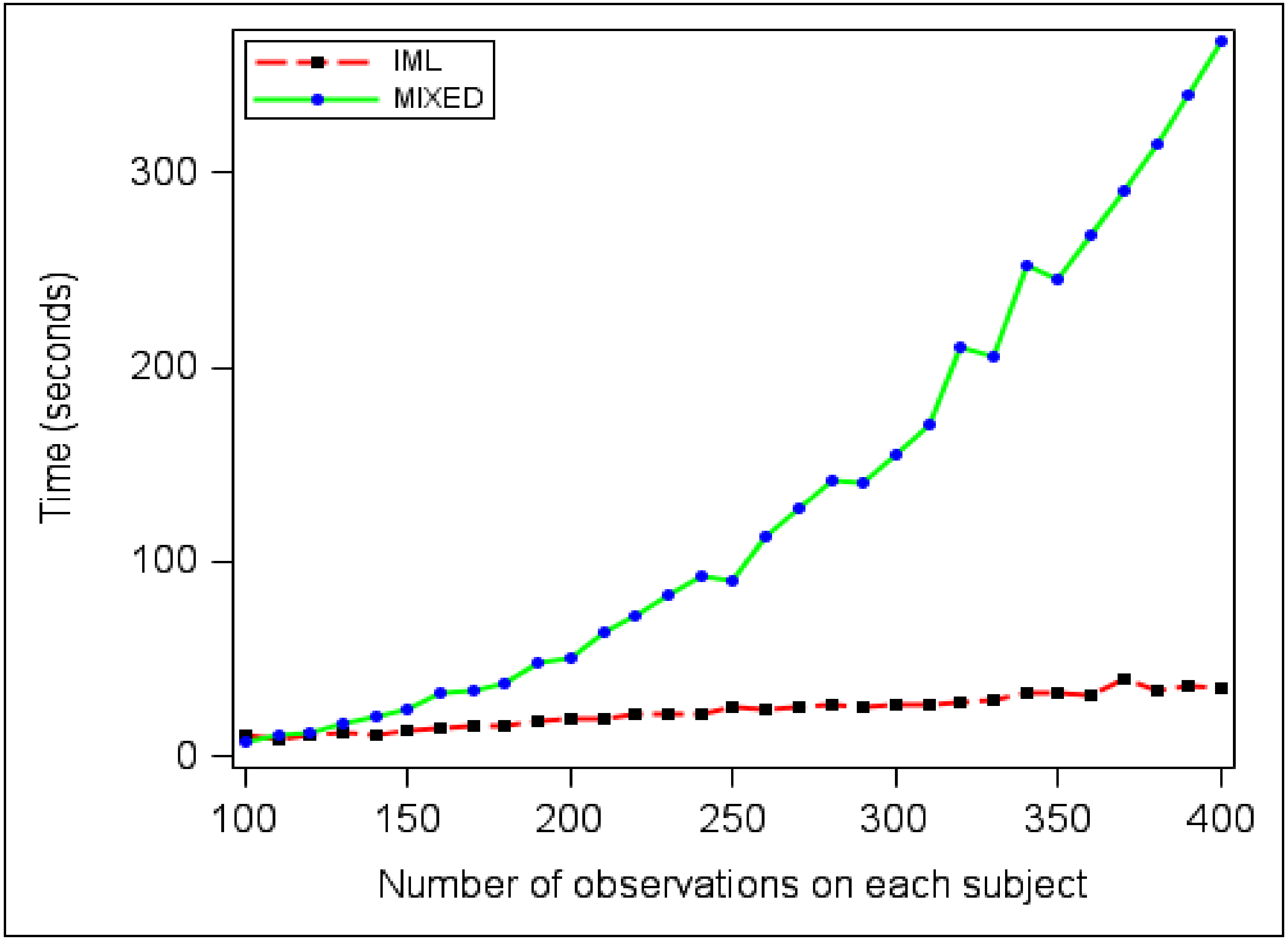
Example 2 results

- 100 subjects, 20 observations, balanced design

Parameter	True value	PROC MIXED		PROC IML	
		Point	SE	Point	SE
ϕ	0.5	.451	.066	.459	.066
μ	1	.969	.112	.971	.112
σ_u^2	0.75	.873	.118	.684	.135
σ_v^2	0.25	.257	.116	.270	.113
τ	1	1.139	.178	1.143	.179
-2*loglik		5904.340		5904.718	

Empirical comparison

- 50 subjects, balanced design



Example 3: multivariate MLM (random intercepts model)

- The combined matrix form:

$$\begin{bmatrix} y1_{ij} \\ y2_{ij} \\ y3_{ij} \end{bmatrix} = \begin{bmatrix} \gamma 1_{00} \\ \gamma 2_{00} \\ \gamma 3_{00} \end{bmatrix} + \begin{bmatrix} u1_{0j} \\ u2_{0j} \\ u3_{0j} \end{bmatrix} + \begin{bmatrix} \varepsilon 1_{ij} \\ \varepsilon 2_{ij} \\ \varepsilon 3_{ij} \end{bmatrix}, \quad \text{where} \begin{bmatrix} \varepsilon 1_{ij} \\ \varepsilon 2_{ij} \\ \varepsilon 3_{ij} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta_{11} & & \\ \delta_{21} & \delta_{22} & \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix} \right).$$

- The random effect vector:

$$\begin{bmatrix} u1_{0j} \\ u2_{0j} \\ u3_{0j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau & & \\ \tau & \tau & \\ \tau & \tau & \tau \end{bmatrix} \right).$$

- Kamata, Bauer, & Miyazaki (2008, Equations 10.8a–10.8g)

State space form

- Re-write the combined matrix form:
- Then we have

$$\begin{bmatrix} y1_{ij} \\ y2_{ij} \\ y3_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u1_{0j} \\ u2_{0j} \\ u3_{0j} \end{bmatrix} + \begin{bmatrix} \gamma 1_{00} \\ \gamma 2_{00} \\ \gamma 3_{00} \end{bmatrix} [1] + \begin{bmatrix} \varepsilon 1_{ij} \\ \varepsilon 2_{ij} \\ \varepsilon 3_{ij} \end{bmatrix}.$$

$$y_{ij} = \begin{bmatrix} y1_{ij} \\ y2_{ij} \\ y3_{ij} \end{bmatrix}, \quad \eta_{ij} = \begin{bmatrix} u1_{0j} \\ u2_{0j} \\ u3_{0j} \end{bmatrix}, \quad x_{ij} = [1], \quad \varepsilon_{ij} = \begin{bmatrix} \varepsilon 1_{ij} \\ \varepsilon 2_{ij} \\ \varepsilon 3_{ij} \end{bmatrix}, \quad \zeta_{ij} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\tau_{\cdot j} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \Lambda_{\cdot j} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Gamma_{\cdot j} = \begin{bmatrix} \gamma 1_{00} \\ \gamma 2_{00} \\ \gamma 3_{00} \end{bmatrix}, \quad \Theta_{\cdot j} = \begin{bmatrix} \delta_{11} & & \\ \delta_{21} & \delta_{22} & \\ \delta_{31} & \delta_{32} & \delta_{33} \end{bmatrix},$$

$$\alpha_{\cdot j} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathrm{B}_{\cdot j} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Upsilon_{\cdot j} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad \Psi_{\cdot j} = \begin{bmatrix} 0 & & \\ 0 & 0 & \\ 0 & 0 & 0 \end{bmatrix}.$$

Initializing the KF

- Initialized with

$$\eta_{0|0,j} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathrm{P}_{0|0,j} = \begin{bmatrix} \tau & & \\ \tau & \tau & \\ \tau & \tau & \tau \end{bmatrix}$$

for all *j*.

Example 3 results

	MLwiN		Mplus		PROC MIXED		PROC IML	
Parameter	Point	SE	Point	SE	Point	SE	Point	SE
$\gamma 1_{00}$	3.677	.072	3.677	.072	3.677	.072	3.677	.072
$\gamma 2_{00}$	3.717	.072	3.717	.072	3.717	.072	3.717	.072
$\gamma 3_{00}$	3.721	.072	3.721	.072	3.721	.072	3.721	.072
τ	.130	.038	.129	.037	.130	.037	.130	.037
δ_{11}	3.036	.070	3.036	.071	3.036	.071	3.036	.071
δ_{12}	1.206	.054	1.206	.054	1.206	.054	1.206	.054
δ_{22}	3.096	.072	3.095	.072	3.096	.072	3.096	.072
δ_{31}	1.328	.054	1.328	.054	1.328	.054	1.328	.054
δ_{32}	1.290	.054	1.290	.054	1.290	.054	1.290	.054
δ_{33}	2.966	.069	2.966	.069	2.966	.069	2.966	.069
-2*loglik	42658.679		42657.388		42658.678		42658.678	

Summary

- The results from the state space approach are identical to those from the MLM and SEM software, except that
- The estimated variance of the innovation in the AR(1) process is biased from both PROC MIXED and PROC IML.
- The state space approach is more efficient because matrix operation is implemented at each observation, instead of at the cluster level.

Discussions

- Software: SAS/IML, R, MATLAB, C++, C, FORTRAN
- Further extension to multilevel CFA and multilevel SEM

Conclusions

- In general, the solution to representing the multilevel models using state space forms is **to include the random effects in the latent state vector**; and then, **initialize the Kalman recursion by properly specifying the latent state vector ($\eta_{0|0,j}$) and its associated covariance matrix ($\mathbf{P}_{0|0,j}$)** to subsume the parameters at the cluster level.