Autoregressive models in the context of structural equation and multilevel modelling

Professor Asko Tolvanen
Methodology Center for Human Sciences
University of Jyväskylä

First, the basic autoregressive model in the SEM and MLM context is presented.

These two basic models produce equal results and points the theoretical differences from which direction models can be develop by using SEM or MLM.

Reading:

Handbook of Structural equation modeling. Edited by Rick H. Hoyle
Introduction to Statistical Mediation Analysis. David P. MacKinnon
Autoregressive model in SEM context

Data for SEM (Structural equation model)

One row for each individual
Repeated measures are named as variables hel1, hel2, hel3,...
Autoregressive SEM model by using observed variables

\[ \text{Tab}_1 = \text{intercept}_1 + \varepsilon_1 \quad \varepsilon_1 \sim N(0, \delta_1^2) \]

\[ \text{Tab}_t = \text{intercept}_2 + \beta \times \text{Tab}_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \delta^2) \]

Tab1 – Tab6: Six repeated measures of task avoidance (sum score which reliability is over .80)

**AR(1) Autoregressive model of first order**

**AR(p) Autoregressive model of p order**

Because there are measurement errors, the estimates can be seriously biased.

---

Interpretation of this model is that the observed value \(\text{TAB}_t\) depends only on the previous value of \(\text{TAB}_{t-1}\) (e.g. two individuals who have equal score at time \(t-1\) has the equal predicted value at time \(t\) independent of previous values at \(t-2\)). This interpretation is totally different than, for example, in growth curve models.

By constraining \(\delta^2 = \beta^2 \times \delta_1^2 + \delta^2\), we suppose that the process is stable. Interpretation is independent of time.
Interpretation of beta coefficient using raw data –

Beta lower than one ($\beta = .707$): individual differences at time $t$ are half of the variance at time $t-1$ (stability $0.707 \times 0.707 = 0.5$) and residual variance $\delta^2$ due to changes between $t-1$ and $t$.

Beta equal to one ($\beta = 1.0$): individual differences at time $t$ are sum of variance at time $t-1$ and residual variance $\delta^2$ due to changes between $t-1$ and $t$. All the individual differences are predicted to appear in next measurement added to the individual changes.

Beta greater than one ($\beta = 2.0$): individual differences at time $t$ are two times the variance at time $t-1$ (stability) and residual variance $\delta^2$ due to changes between $t-1$ and $t$.

\[
\begin{align*}
\text{Intercepts} & \\
\text{TAB2} & = 1.898 \\
\text{TAB3} & = 1.898 \\
\text{TAB4} & = 1.898 \\
\text{TAB5} & = 1.898 \\
\text{TAB6} & = 1.898 \\
\text{Variances} & \\
\text{TAB1} & = 11.529 \\
\text{TAB2} & = 8.990 \\
\text{TAB3} & = 8.990 \\
\text{TAB4} & = 8.990 \\
\text{TAB5} & = 8.990 \\
\text{TAB6} & = 8.990
\end{align*}
\]

**Total variance increases slightly because the equality constraint is not done for first measurement.**

About 30% of the total variance are related to preceding measurement.

About 70% of the total variance is related to changes between successive measurement and partly to the measurement error.
Autoregressive SEM by using observed variables

Chi-Square Test of Model Fit
Value                            109.183*
Degrees of Freedom                    22
P-Value                           0.0000
Scaling Correction Factor         1.0991
for MLR

RMSEA (Root Mean Square Error Of Approximation)
Estimate                           0.130
90 Percent C.I.                    0.106  0.154
Probability RMSEA <= .05           0.000

CFI/TLI
CFI                                0.422
TLI                                0.606

SRMR (Standardized Root Mean Square Residual)
Value                              0.170

Model fit is poor!
The model cannot explain the associations between observed variables!
One reason to bad fit could be that the reliability is lower than 1.

For good fitting model

One reason to bad fit could be that the reliability is lower than 1.

Autoregressive SEM by using observed variables

MODEL MODIFICATION INDICES
Minimum M.I. value for printing the modification index 8.000

M.I. E.P.C. Std E.P.C. Std YX E.P.C.

ON Statements
TAB3 ON TAB3 16.320 0.188 0.188 0.188
TAB3 ON TAB4 20.455 0.256 0.256 0.256
TAB3 ON TAB5 9.810 0.173 0.173 0.173
TAB3 ON TAB6 28.512 0.290 0.290 0.290
TAB3 ON TAB1 24.854 0.249 0.249 0.267
TAB4 ON TAB5 8.735 -0.165 -0.165 -0.165
TAB5 ON TAB4 8.042 -0.129 -0.129 -0.129
TAB6 ON TAB3 10.465 0.171 0.171 0.171

WITH Statements
TAB5 WITH TAB4 9.890 -2.225 -2.225 -0.248
TAB6 WITH TAB3 10.568 2.110 2.110 0.235
TAB1 WITH TAB4 15.730 3.021 3.021 0.297
TAB1 WITH TAB6 9.648 2.358 2.358 0.232

Means/Intercepts/Thresholds
[ TAB3 ] 20.715 0.886 0.886 0.281

Lower bounds on sample size in structural equation modeling
J. Christopher Westland
Advantage of the autoregressive SEM model is that we could test the stability hypotheses by using overall fit indices.

Further, if the stability hypotheses are not supported, we could respecify the model with the help of modification indices.

When the model complexity increases, the fit indices need larger sample size to work well.

(Bentler: 5 times the estimated parameters?)
(Kline: 20 times the number of estimated parameters)

When modifying the model, the fit indices work more as descriptive purposes.

When constraining some parameters, the modification index are not computed

Autoregressive model in Multilevel context
Successive measurements are named

hel1, anx1, tab1, math1 having values in time=1,2,...,6

hel2, anx2, tab2, math2 having values in time=1,2,...,5

which are actually values in time 2,3,...,6

Autoregressive MLM by using observed variables

\[
\begin{align*}
T_{ab2} &= \text{intercept}_j + \beta \times T_{ab1} + \epsilon \\
\epsilon &\sim N(0, \delta^2) \\
\text{intercept}_j &= \alpha + \mu_j \\
\mu_j &\sim N(0, 0)
\end{align*}
\]

Tab1 is not group centered but defined as WITHIN variable. Intercepts are equal across measurements.

--> Very similar model than the SEM model presented before.
### MODEL RESULTS

**Within Level**

<table>
<thead>
<tr>
<th>TAB2 ON TAB1</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.312</td>
<td>0.040</td>
<td>7.879</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.319</td>
<td>0.039</td>
<td>7.943</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Residual Variances**

<table>
<thead>
<tr>
<th>TAB2</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.008</td>
<td>0.512</td>
<td>17.586</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>8.990</td>
<td>0.496</td>
<td>18.113</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Between Level**

**Means**

<table>
<thead>
<tr>
<th>TAB2</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.878</td>
<td>0.124</td>
<td>15.195</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>1.898</td>
<td>0.122</td>
<td>15.513</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Variances**

<table>
<thead>
<tr>
<th>TAB2</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.00</td>
</tr>
</tbody>
</table>

---

### Autoregressive MLM by using observed variables

**WITHIN**

\[ \beta \]

\[ \text{Tab1} \rightarrow \beta \rightarrow \text{Tab2} \rightarrow \varepsilon \]

**BETWEEN**

\[ \alpha \]

\[ \Delta \rightarrow 0 \rightarrow \text{Tab2} \]

---

**Chi-Square Test of Model Fit**

- **Value**: 15.785
- **Degrees of Freedom**: 1
- **P-Value**: 0.0001
- **Scaling Correction Factor**: 0.8347 for MLR

**RMSEA (Root Mean Square Error Of Approximation)**

- **Estimate**: 0.119

**CFI/TLI**

- **CFI**: 0.000
- **TLI**: -0.264

**SRMR (Standardized Root Mean Square Residual)**

- **Value for Within**: 0.072
- **Value for Between**: 0.000
### Autoregressive MLM by using observed variables

**WITHIN**

\[
\begin{align*}
\text{Tab1} & \quad \beta \quad \text{Intercept} \\
& \quad \rightarrow \\
\text{Tab2} & \quad \leftarrow \varepsilon
\end{align*}
\]

**BETWEEN**

\[
\begin{align*}
\alpha & \quad \text{Tab2} \\
& \quad \rightarrow \\
? & \quad \rightarrow \text{Tab2}
\end{align*}
\]

### MODEL MODIFICATION INDICES

- Minimum M.I. value for printing the modification index: 4.000

<table>
<thead>
<tr>
<th>M.I.</th>
<th>E.P.C.</th>
<th>Std E.P.C.</th>
<th>StdYX</th>
<th>E.P.C.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Within Level**

**Between Level**

### Variances/Residual Variances

<table>
<thead>
<tr>
<th>Variances/Residual Variances</th>
<th>Tab2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.492</td>
<td>0.430</td>
<td>0.430</td>
<td>4297.974</td>
</tr>
</tbody>
</table>

### Autoregressive MLM by using observed variables

**WITHIN**

\[
\begin{align*}
\text{Tab1} & \quad \beta \quad \text{Intercept} \\
& \quad \rightarrow \\
\text{Tab2} & \quad \leftarrow \varepsilon
\end{align*}
\]

**BETWEEN**

\[
\begin{align*}
\alpha & \quad \text{Tab2} \\
& \quad \rightarrow \\
? & \quad \rightarrow \text{Tab2}
\end{align*}
\]

\[
\begin{align*}
\text{Tab2}_j = \text{intercept}_j + \beta \times \text{Tab1} + \varepsilon & \quad \varepsilon \sim N(0, \delta_\varepsilon^2) \\
\text{intercept}_j = \alpha + \mu_j & \quad \mu_j \sim N(0, \delta_\mu^2)
\end{align*}
\]
Autoregressive MLM: Between level variance estimated freely
TAB1 is not group centered!

<table>
<thead>
<tr>
<th>MODEL RESULTS</th>
<th>Two-Tailed</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2 ON TAB1</td>
<td></td>
<td>0.172</td>
<td>0.053</td>
<td>3.228</td>
<td>0.001</td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
<td>0.312</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>7.828</td>
<td>0.537</td>
<td>14.570</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Between Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>2.277</td>
<td>0.182</td>
<td>12.477</td>
<td>0.000</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>1.382</td>
<td>0.533</td>
<td>2.596</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Autoregressive MLM: Between level variance estimated freely
TAB1 is group centered!

Results for model in which TAB1 is not group centered are colored red

<table>
<thead>
<tr>
<th>MODEL RESULTS</th>
<th>Two-Tailed</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2 ON TAB1</td>
<td></td>
<td>-0.117</td>
<td>0.032</td>
<td>-3.637</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>7.199</td>
<td>0.453</td>
<td>15.884</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Between Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>2.771</td>
<td>0.139</td>
<td>19.984</td>
<td>0.000</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB2</td>
<td></td>
<td>2.741</td>
<td>0.529</td>
<td>5.179</td>
<td>0.000</td>
</tr>
</tbody>
</table>
The advantage of ML model is that it can separate typical individual level variation and further model the variability in process around the individual level.

The structure between repeated measures cannot test!

It is obvious that we need for theoretical knowledge to specify the model which further guides as to use SEM or MLM!
Modelling using SEM

Autoregressive SEM by using observed variables

Chisquare Test of Model Fit
Value 19.104
Degrees of Freedom 15
P-Value 0.1885
RMSEA (Root Mean Square Error Of Approximation)
Estimate 0.036
90 Percent C.I. 0.000-0.076
Probability RMSEA <= .05 0.674
CFI/TLI
CFI 0.976
TLI 0.976
SRMR (Standardized Root Mean Square Residual)
Value 0.047

Model is identified (possible to estimate) if for example equality constraint is done for some of the error variances.

Advantage: The residuals (measurement error) can be estimated -> reliability estimates

\[ \beta = 0.87 \]
\[ \beta_{54} = 0.28 \]
\[ \beta_{62} = 0.34 \]
It is possible to test the need for latent factor, by capturing individual level variation.

These two models are equivalent producing equal fit to the data.

$\beta_1 = 1 + \beta_2$
Crosslagged autoregressive SEM model

Task avoidance predicts the math skills in next measurement.

Anxiety predicts task avoidance and task avoidance predicts math.
Task avoidance can be interpreted as mediator between anxiety and math.

Limitation: There are other models that can explain associations which are shown in this model – for example we have some effect outside of this model.
By using latent factors in longitudinal data, it is possible to isolate the measurement error by using single item. When using sum score instead of items, model complexity decreases and we are maybe able to use fit indices testing the hypothetical model.

By using sum score instead of CFA, the are some requirements related to measurement structure:

- Invariance of factor loadings across measurements holds
- Invariance of intercepts holds
- Invariance of error variances holds

Each item should measure the same construct and the variance of item is due to construct and measurement error.

\[
\text{SUM1} = \text{Time1} + \text{Time2} + \text{Time3}
\]

Modelling with MLM
Crosslagged autoregressive MLM model \textit{(Math1 and Tab1 defined as within variables)}

\begin{itemize}
  \item \textbf{WITHIN} \ 
  \begin{align*}
  \text{Var(Math1)} &= 202.136 \\
  \text{Var(Tab1)} &= 10.521 \\
  \end{align*} \\
  \begin{align*}
  \text{Math1} &\to 0.967 \to \text{Math2} \to 0.7846 \to \text{Tab2} \\
  \text{Tab1} &\to 0.173 \to \text{Tab2} \\
  \end{align*} \\
  \textbf{BETWEEN} \ 
  \begin{align*}
  \text{WITHIN} &\to 0.028 \text{ n.s.} \to \text{BETWEEN} \to 1.350 \text{ MLM} \\
  \text{df} &= 0 \\
  \end{align*} \\
\end{itemize}

Task avoidance predicts the math skills in next measurement.

Crosslagged autoregressive MLM model \textit{(Math1 and Tab1 group mean centered and defined as within variables)}

\begin{itemize}
  \item \textbf{WITHIN} \ 
  \begin{align*}
  \text{Var(Math1)} &= 170.874 \\
  \text{Var(Tab1)} &= 5.961 \\
  \end{align*} \\
  \begin{align*}
  \text{Math1} &\to 0.930 \to \text{Math2} \to -4.641 \to \text{Tab2} \\
  \text{Tab1} &\to -0.116 \to \text{Tab2} \\
  \end{align*} \\
  \textbf{BETWEEN} \ 
  \begin{align*}
  \text{WITHIN} &\to 0.132 \text{ p= 0.055} \to \text{BETWEEN} \text{ MLM} \\
  \text{df} &= 0 \\
  \end{align*} \\
\end{itemize}

Task avoidance predicts the math skills in next measurement (p=.066).
Crosslagged autoregressive MLM model

WITHIN

```
+-------------------+-------------------+
| Math1             | Math2             |
| 282.549           | 26.056            |
| 0.965             | 0.113             |
| 0.383             | 0.345             |
| 6.843             | 6.496             |
```

BETWEEN

```
+-------------------+-------------------+
| Math1             | Math2             |
| 0.999             | STDYX             |
| -0.64 - -0.67      |
```

Also further advantage of the ML model is that the variation in the regression coefficient can be estimated and predicted. (Multilevel moderator model)
Within

\[ i_M \rightarrow i_Y \]

Between

\[ X \rightarrow i_Y \]

\[ X \rightarrow i_M \]

Multilevel mediator model!

If the interval between repeated measures are not equal but all Individuals are measured in certain time points.

In SEM we can use constraint \[ \beta_2 = \beta_1^2 \]

\[ \eta_t = \beta^t \times \eta_{t-1} \]

Constraint is not possible in multilevel model?
Combining SEM and MLM

WITHIN

\[ \lambda_1 \]
\[ \lambda_2 \]
\[ 1 \]

\[ \text{ANX} \]
\[ \text{TAB} \]
\[ \text{HEL} \]
\[ \text{Sw} \]
\[ \text{Math} \]

BETWEEN

\[ \lambda_1 \]
\[ \lambda_2 \]
\[ 1 \]

\[ \text{ANX} \]
\[ \text{TAB} \]
\[ \text{HEL} \]
\[ \text{Sb} \]
\[ \text{Math} \]
Thank you!