Mplus 8: Dynamic SEM

Time series analysis and state-space modeling

Ellen L. Hamaker
Utrecht University

Tihomir Asparouhov & Bengt Muthén
Muthén & Muthén

May 23, 2016
Cattell’s data box
Cross-sectional research: A single snapshot
Panel research: A few snapshots
Time series data: Looking at the movie
Time series data: Looking at the movie
Outline

- Why time series analysis?
- Autocorrelation
- ARMA models
- Stationarity
- The state-space model
- Kalman filter and parameter estimation
- Diverse models in state-space format
- Miscellaneous
What is time series analysis?

Time series analysis is a class of techniques that is used in econometrics, seismology, meteorology, control engineering, and signal processing.

Main characteristics:

- \( N=1 \) technique
- \( T \) is large (say \( >50 \))
- concerned with *trends*, *cycles* and *autocorrelation structure* (i.e., serial dependency)
- goal: forecasting (\( \neq \) prediction)
TSA in the social and medical sciences

In **sociology**:  
- quarterly unemployment numbers  
- effect of alcohol consumption per capita on criminal violence rates  
- effect of suicide news on suicide rates

In **medical research**:  
- effect of safety warnings on antidepressants use  
- effects of pain control strategies  
- effect of 9/11 attacks on weekly psychiatric patient admissions

In **psychology**:  
- network of symptoms in depressive patient  
- effect of feedback on academic performance  
- effect of an intervention on the relationship between stress and affect
Intensive longitudinal data are gathered using:

- daily diary with end-of-day-measurements (self-report)
- experience sampling method (self-report)
- ecological momentary assessment (self-report)
- ambulatory assessment (including physiological variables)
- event contingency (self-report)
- observational measurements (expert rater)

For more info on methodology, check out:

- Tamlin Conner (e.g., her seminar with Joshua Smyth on YouTube)
- Society for Ambulatory Assessment
- Trull and Ebner-Priemer (2013)
It’s a revolution!

Publications on experience sampling, ambulatory assessment, ecological momentary assessment, or daily diary

Year

Number of publications

PsycINFO

PubMed
A fundamental problem in a nutshell

Cross-sectional relationship

Within-person relationship

Between-person relationship

Taken from Hamaker (2012).
Three perspectives on data

Cross-sectional

Within

Between

Taken from Hamaker (2012).
Interindividual differences in intraindividual variation

Taken from Hamaker and Grasman (2014).
Cross-sectional correlations: A blend

Schmitz (2000):

\[ r_{cs} = \eta^2 r_b + (1 - \eta^2) r_w \]

where

- \( r_{cs} \) is the cross-sectional correlation
- \( r_b \) is the between-person correlation
- \( r_w \) is the within-person correlation
- \( \eta^2 \) is the proportion of between-person variance of the total variability

Consequences:

- Cross-sectional and panel research may result in an “uninterpretable blend” of within-person and between-person relationships (cf. Raudenbush and Bryk, 2002)
- In N=1 time series analysis there is only within-person variance
Outline

• Why time series analysis?
• **Autocorrelation**
• ARMA models
• Stationarity
• The state-space model
• Kalman filter and parameter estimation
• Diverse models in state-space format
• Miscellaneous
### Lags

<table>
<thead>
<tr>
<th>Y</th>
<th>Y at lag 1</th>
<th>Y at lag 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>$y_1$</td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>$y_2$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>$y_4$</td>
<td>$y_3$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>$y_5$</td>
<td>$y_4$</td>
<td>$y_3$</td>
</tr>
<tr>
<td>$y_6$</td>
<td>$y_5$</td>
<td>$y_4$</td>
</tr>
<tr>
<td>$y_7$</td>
<td>$y_6$</td>
<td>$y_5$</td>
</tr>
<tr>
<td>$y_8$</td>
<td>$y_7$</td>
<td>$y_6$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$y_T$</td>
<td>$y_{T-1}$</td>
<td>$y_{T-2}$</td>
</tr>
<tr>
<td>$y_T$</td>
<td>$y_{T-1}$</td>
<td>$y_{T-1}$</td>
</tr>
<tr>
<td>$y_T$</td>
<td></td>
<td>$y_T$</td>
</tr>
</tbody>
</table>
Autocorrelation function (ACF)

The ACF and the PACF can be used as diagnostic tools to determine the nature of the underlying process.

**Variance (or: auto-covariance at lag 0):**

\[ \gamma_0 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \bar{y}_t)^2 \]

**Auto-covariance at lag \( k \):**

\[ \gamma_k = \frac{1}{T-k} \sum_{t=k+1}^{T} (y_t - \bar{y}_t)(y_{t-k} - \bar{y}_t) \]

**Autocorrelation at lag \( k \):**

\[ \rho_k = \frac{\gamma_k}{\gamma_0} \]
Partial autocorrelation function (PACF)

Partial autocorrelation at lag $k$ is the correlation between $y_t$ and $y_{t-k}$ after removing the effect of the intermediate observations (i.e., $y_{t-1}$ to $y_{t-k+1}$).

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$Y$ at lag 1</th>
<th>$Y$ at lag 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_2$</td>
<td>$y_1$</td>
<td></td>
</tr>
<tr>
<td>$y_3$</td>
<td>$y_2$</td>
<td>$y_1$</td>
</tr>
<tr>
<td>$y_4$</td>
<td>$y_3$</td>
<td>$y_2$</td>
</tr>
<tr>
<td>$y_5$</td>
<td>$y_4$</td>
<td>$y_3$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$y_T$</td>
<td>$y_{T-1}$</td>
<td>$y_{T-2}$</td>
</tr>
<tr>
<td></td>
<td>$y_T$</td>
<td>$y_{T-1}$</td>
</tr>
<tr>
<td></td>
<td>$y_T$</td>
<td>$y_T$</td>
</tr>
</tbody>
</table>

For instance: Is there a relationship between yesterday’s positive affect and tomorrow’s positive affect above and beyond their relationship to today’s positive affect?
Sequence, ACF and PACF

White Noise process

First-order AR process

Second-order AR process
Outline

• Why time series analysis?
• Autocorrelation
• **ARMA models**
• Stationarity
• The state-space model
• Kalman filter and parameter estimation
• Diverse models in state-space format
• Miscellaneous
**AR(1):** \( y_t = \phi_1 y_{t-1} + u_t \)

Example with \( \phi_1 = 0.7 \) and \( \phi_1 = -0.7 \):
AR(2): \[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t \]

Example with \( \phi_1 = 1.2 \) and \( \phi_2 = -0.7 \) and with \( \phi_1 = 0.2 \) and \( \phi_2 = 0.7 \):
MA(1): \[ y_t = u_t - \theta_1 u_{t-1} \]

Examples with \( \theta_1 = 0.7 \) and with \( \theta_1 = -0.7 \):
MA(2): \[ y_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \]

Examples with \( \theta_1 = 1.2 \) and \( \theta_2 = -0.7 \), and with \( \theta_1 = 0.2 \) and \( \theta_2 = 0.7 \):
**ARMA(1,1):** \[ y_t = \phi y_{t-1} + u_t - \theta_1 u_{t-1} \]

Example with \( \phi_1 = 0.8 \) and \( \theta_1 = 0.8 \), and with \( \phi_1 = -0.8 \) and \( \theta_1 = -0.8 \):
Pure AR, pure MA, or an ARMA($p, q$)?

In general:

- an AR($p$) can **always** be written as an MA($\infty$)
- an MA($q$) can **always** be written as an AR($\infty$)

Other (rather unexpected) results found by **Granger and Morris (1976)**:

- AR(1) + WN $\rightarrow$ ARMA(1,1)
- AR(1) + AR(1) $\rightarrow$ ARMA(2,1)
- MA(1) + WN $\rightarrow$ MA(1)

You may consider:

- interpretation (social sciences)
- forecasting (econometrics)
- parsimony
Outline

• Why time series analysis?
• Autocorrelation
• ARMA models
• **Stationarity**
• The state-space model
• Kalman filter and parameter estimation
• Diverse models in state-space format
• Miscellaneous
Stationarity

Stationarity is an important concept in time series analysis:

- is based on using backshift operators and the unit root circle (as all introductory texts on time series analysis do!)
- implies that all moments (i.e., means, variances, covariances, lagged covariances, etc.) are independent of time

For instance:

- mean is constant over time
- $\gamma_k$ depends on the lag $k$, not on $t$ (i.e., the occasion itself)

Two typical examples of nonstationary processes:

- trends over time (including cycles?)
- random walk: $y_t = y_t + e_t$
Stationarity of an AR(p)

For an AR(1) to be stationary, $|\phi| < 1$.

For an AR(2) to be stationary we need:

- $\phi_2 - \phi_1 < 1$
- $\phi_2 + \phi_1 < 1$
- $|\phi_2| < 1$

which leads to the following triangle:

(Check out: http://freakonometrics.hypotheses.org/12081)
Outline

• Why time series analysis?
• Autocorrelation
• ARMA models
• Stationarity
• **The state-space model**
• Kalman filter and parameter estimation
• Diverse models in state-space format
• Miscellaneous
Rocket science

State space model with **known parameters**:

- **Kalman filter** predicts the future state (e.g., the location of your space rocket), based on current and previous observations (on-line procedure)
- **Kalman smoother** predicts the state based on previous, current and future observations (off-line procedure)

Often, the **parameter values are NOT known**.

Then, certain **by-products** of the Kalman filter/smoother can be used in a **likelihood function** (see later).
The basic framework

**Measurement equation**

\[ y_t = c_t + Z_t a_t + e_t \quad \text{with} \quad e_t \sim \text{MN}(0, GG_t) \]

- \( c_t \) is the vector with intercepts in the measurement equation
- \( Z_t \) is the matrix with factor loadings
- \( GG_t \) is the covariance matrix of the measurement errors

**Transition equation**

\[ a_t = d_t + T_t a_{t-1} + u_t \quad \text{with} \quad u_t \sim \text{MN}(0, HH_t) \]

- \( d_t \) is the vector with intercepts in the transition equation
- \( T_t \) is the matrix with cross- and auto-regressive coefficients
- \( HH_t \) is the covariance matrix of the dynamic errors

In a more basic version these model matrices are fixed over time.
Measurement equation: regressing $y_t$ on $a_t$

$y_t = c + Z a_t + e_t$

$e_t \sim MN(0, \mathbb{G})$
Transition equation: Regressing $a_t$ on $a_{t-1}$

\[ a_t = d + T a_{t-1} + u_t \]

\[ u_t \sim MN(0, HH) \]
State-space model = Latent VAR(1) model

\[ y_t = c + Z a_t + e_t \]
\[ e_t \sim MN(0, GG) \]

\[ a_t = d + T a_{t-1} + u_t \]
\[ u_t \sim MN(0, HH) \]
Outline

- Why time series analysis?
- Autocorrelation
- ARMA models
- Stationarity
- The state-space model
- Kalman filter and parameter estimation
- Diverse univariate models in state-space format
- Miscellaneous
State-space model versus SEM

Two ways in which SEM can be use to do TSA:

**Toeplitz method**, based on making lagged variables

\[
\begin{array}{cccc}
  y_1 & y_1 & \ldots & y_T \\
  y_2 & y_2 & \ldots & y_{T-1} \\
  y_3 & y_3 & \ldots & \\
  y_4 & y_3 & \vdots & \\
  \vdots & \vdots & \ddots & \\
  y_T & y_{T-1} & \ldots & y_1 \\
\end{array}
\]

**Advantage**: easy

**Disadvantage**: violates assumption of independent cases (=rows); no true ML estimates (and wrong fit measures)

(cf. Hamaker, Dolan & Molenaar, 2002)

**Raw maximum likelihood estimation**, based on N=1

\[
\begin{array}{cccc}
  y_1 & y_2 & y_3 & \ldots & y_T \\
\end{array}
\]

**Advantage**: gives ML estimates

**Disadvantage**: requires inversion of (at least) a $T \times T$ matrix (computationally troublesome)

(cf. Hamaker, Dolan and Molenaar, 2003)

See Chow, Ho, Hamaker and Dolan (2010) for further comparison of state-space modeling and SEM.
Kalman filter for parameter estimation

The Kalman filter can be used to predict future states when the parameters are known.

In practice, the parameter values are often unknown.

In that case, by-products of the Kalman filter can be used to estimate the parameters:

- the one-step-ahead-prediction error $e_{t|t-1} = y_t - y_{t|t-1}$
- the covariance matrix of $e_{t|t-1}$ (i.e., $F_t$)

These are plugged into a likelihood function, which is then optimized with respect to the unknown parameters.

Hence, for each set of possible parameter values, the entire Kalman filter is run from $t = 1$ to $t = T$. 
Kalman filter for parameter estimation

Choose $\hat{\theta}^{(i)}$ and let $L(\hat{\theta}^{(i)}|0) = 1$

Choose $a_{0|0}$ and $P_{0|0}$

Predict $a_t$ and $P_t$:

- $a_{\eta|t-1} = d + Ta_{\eta|t-1}$
- $P_{\eta|t-1} = TP_{\eta|t-1}T' + HH'$

Predict $y_t$ and compute $F_t$:

- $y_{\eta|t-1} = c + Za_{\eta|t-1}$
- $F_t = ZP_{\eta|t-1}Z' + GG'$

Update prediction of $a_t$ and $P_t$

- $a_{\eta|t} = a_{\eta|t-1} + P_{\eta|t-1}Z'F_t^{-1}e_{\eta|t-1}$
- $P_{\eta|t} = P_{\eta|t-1} - P_{\eta|t-1}Z'F_t^{-1}ZP_{\eta|t-1}$

Compute one-step-ahead prediction error:

- $e_{\eta|t-1} = y_t - y_{\eta|t-1}$

Enter $e_{\eta|t-1}$ in the likelihood function of occasion $t$:

$$L(\hat{\theta}^{(i)}|t) = L(\hat{\theta}^{(i)}|t-1) \left( 2\pi \right)^{-NY/2} \det(F_t)^{-1/2} \exp \left\{ -\frac{1}{2} e'_{\eta|t-1}F_t^{-1}e_{\eta|t-1} \right\}$$

$t = T$?

YES

NO, $t = t + 1$
Outline

• Why time series analysis?
• Autocorrelation
• ARMA models
• Stationarity
• The state-space model
• Kalman filter and parameter estimation
• Diverse models in state-space format
• Miscellaneous
Just a latent vector AR(1) model?

At first sight the state-space model seems to be just a latent VAR(1) model.

However, it is actually a very flexible framework for all sorts of time series models:

- all ARIMA models
- multivariate extensions
- dynamic factor analysis

Extensions may consist of:

- predictors (e.g., time, intervention, weather conditions) in the measurement and/or transition equation
- time-varying parameters
- regime switches (through combination with a hidden Markov process)
**AR(1) in state-space format**

**Measurement equation:**

\[ y_t = c + a_t \]

- \( c \) is a vector containing the unknown mean
- \( Z \) is a 1 by 1 matrix containing 1
- \( GG \) is a zero matrix

**Transition equation:**

\[ a_t = Ta_{t-1} + u_t \]

- \( d \) is a zero vector
- \( T \) is a 1 by 1 matrix containing the autoregressive parameter
- \( HH \) is a 1 by 1 covariance matrix containing the variance of the innovations
**AR(1) with measurement error**

### Measurement equation:

\[ y_t = c + a_t + e_t \]

- \( c \) is a vector containing the unknown mean
- \( Z \) is a 1 by 1 matrix containing 1
- \( GG \) is a 1 by 1 covariance matrix with the variance of the measurement error

### Transition equation:

\[ a_t = Ta_{t-1} + u_t \]

- \( d \) is a zero vector
- \( T \) is a 1 by 1 matrix containing the autoregressive parameter
- \( HH \) is a 1 by 1 covariance matrix containing the variance of the innovations
AR(2) in state-space format

**Measurement equation:**

\[ y_t = c + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} a_t \\ a_{t-1} \end{bmatrix} = c + a_t \]

where \( GG \) is a zero matrix.

**Transition equation:**

\[
\begin{bmatrix} a_t \\ a_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-2} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \end{bmatrix} = \begin{bmatrix} \phi_1 a_{t-1} + \phi_2 a_{t-2} + u_t \\ a_{t-1} \end{bmatrix}
\]

- \( d \) is a zero vector
- \( HH \) is a 2 by 2 covariance matrix containing only the variance of the innovations (element 1,1)
Bivariate VAR(1) in state-space format

**The measurement equation:**

\[
\begin{bmatrix}
y_{1,t} \\
y_{2,t}
\end{bmatrix} =
\begin{bmatrix}
c_1 \\
c_2
\end{bmatrix} +
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
a_{1,t} \\
a_{2,t}
\end{bmatrix}
\]

where \(GG\) is a zero matrix.

**The transition equation:**

\[
\begin{bmatrix}
a_{1,t} \\
a_{2,t}
\end{bmatrix} =
\begin{bmatrix}
\phi_{11} & \phi_{21} \\
\phi_{12} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
a_{1,t-1} \\
a_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
u_{1,t} \\
u_{2,t}
\end{bmatrix} =
\begin{bmatrix}
\phi_{11} a_{1,t-1} + \phi_{21} a_{2,t-1} + u_{1,t} \\
\phi_{12} a_{1,t-1} + \phi_{22} a_{2,t-1} + u_{2,t}
\end{bmatrix}
\]

- \(d\) is a zero vector
- \(HH\) is a 2 by 2 covariance matrix of \(u_{1,t}\) and \(u_{2,t}\)
Graphical representations

\[ \begin{align*}
\text{NE}_{t-1} & \to \text{NE}_t & \to \text{NE}_{t+1} \\
\text{Ru}_{t-1} & \to \text{Ru}_t & \to \text{Ru}_{t+1} \\
\text{S}_{t-1} & \to \text{S}_t & \to \text{S}_{t+1} \\
\text{NA}_{t-1} & \to \text{NA}_t & \to \text{NA}_{t+1}
\end{align*} \]
Applications of VAR models

VAR models are of interest, because

- they allow you to study **Granger causality**: Can you predict $Y$ from $X$, after controlling for previous levels of $Y$?
- they allow you to determine which variable is "**causally dominant**" when there are reciprocal effects
- they can be interpreted as **networks** (alternative to latent variable approach)

Some interesting **replicated VAR applications**

- Schmitz and Skinner (1994): Perceived control, effort and academic performance
- Rosmalen et al. (2012): Depression and physical activity
- Snippe et al. (2014): Mindfulness, repetitive thinking and depressive symptoms
- Van Gils et al. (2014): Stress and functional somatic symptoms

In all these studies they find **important differences across individuals.**
Dynamic factor model

Dynamic factor analysis is used for time series data consisting of multiple indicators of an underlying construct.

There are two popular versions:

- at the latent level there is a **VARMA model**; the factor loadings only appear at lag 0
- at the latent level there is **white noise**; the factor loadings appear at different lags (e.g., EEG data)
Outline

• Why time series analysis?
• Autocorrelation
• ARMA models
• Stationarity
• The state-space model
• Kalman filter and parameter estimation
• Diverse univariate models in state-space format
• Miscellaneous
Covariance matrix of the series

For a **univariate AR(1)**, we have: \( \sigma_y^2 = \frac{\sigma_u^2}{1-\phi^2} \).

Similarly, for a **(latent) VAR model** we can express the **covariance matrix** of \( y_t \) in terms of

- lagged regression parameters \( \Phi \)
- covariance matrix of the innovations \( \Gamma \) (i.e., HH in the state-space model)

Specifically (from Kim and Nelson, 1999):

\[
\Sigma_y = mat \left[ (I - \Phi \otimes \Phi)^{-1} \text{vec}(\Gamma) \right]
\]

where

- \( \text{vec()} \) implies you put all the **matrix elements in a vector**
- \( \text{mat()} \) implies you place all the **vector elements in a square matrix**
Model fit

Despite the **similar appearance**, state-space modeling and SEM are **not the same**: For a time series there is no **saturated model** against which we can test other models.

We can **compare our model to other models**, including the white noise model (independence model), using

- log likelihood ratio test (for nested models)
- AIC, BIC, DIC, etc. (for all models)

Fit may be **less interesting** to econometricians and meteorologists: Their primary interest is **forecasting**.
To conclude

- time series analysis is a large class of diverse techniques to analyze N=1 data
- ARMA models are only a small (but basic) part of this
- time series models may be extended with cycles or trends over time
- in psychology we typically have N>1; there are different ways of handling this
References and suggested readings


References and suggested readings


