Box 1 - Example DFA workflow

The general idea is to explain temporal variation in a set of n observed time series using linear combinations of a set of m hidden random walks, where $m << n$. A DFA model has the following structure:

$$\begin{gathered}\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t \text{ where } \mathbf{w}_t \sim \text{MVN}(0, \mathbf{Q}) \\ \mathbf{y}_t = \mathbf{Z}\mathbf{x}_t + \mathbf{a} + \mathbf{v}_t \text{ where } \mathbf{v}_t \sim \text{MVN}(0, \mathbf{R})\end{gathered}.$$

Imagine a case where we had a data set with five observed time series ($n = 5$) and we want to fit a model with three hidden processes ($m = 3$). If we write out our DFA model in matrix form, the process model would look like this:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{t-1} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t$$

And the observation equation would look like

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}_t = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \\ z_{51} & z_{52} & z_{53} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}_t .$$

The process errors would be

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix} \right).$$

And the observation errors would be

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}_t \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{12} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{13} & r_{23} & r_{33} & r_{34} & r_{35} \\ r_{14} & r_{24} & r_{34} & r_{44} & r_{45} \\ r_{15} & r_{25} & r_{35} & r_{45} & r_{55} \end{bmatrix} \right)$$

Constraining a DFA model

If $\mathbf{Z}$, $\mathbf{a}$, and $\mathbf{Q}$ are not constrained, however, then the DFA model above is unidentifiable. Nevertheless, we can use the following parameter constraints to make the model identifiable:

- in the first $m - 1$ rows of $\mathbf{Z}$, the z-value in the j-th column and i-th row is set to zero if $j > i$;
- $\mathbf{a}$ is constrained so that the first m values are set to zero; and
- $\mathbf{Q}$ is set equal to the identity matrix ($\mathbf{I_m}$).

Using these constraints, the process equation for the DFA model above becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{t-1} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t ,$$

and the observation equation becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}_t = \begin{bmatrix} z_{11} & 0 & 0 \\ z_{21} & z_{22} & 0 \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \\ z_{51} & z_{52} & z_{53} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}_t .$$

The process errors would then become

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

but the observation errors would stay the same, such that

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}_t \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{12} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{13} & r_{23} & r_{33} & r_{34} & r_{35} \\ r_{14} & r_{24} & r_{34} & r_{44} & r_{45} \\ r_{15} & r_{25} & r_{35} & r_{45} & r_{55} \end{bmatrix} \right).$$

Different error structures

The example observation equation we used above had what we refer to as an "unconstrained" variance-covariance matrix $\mathbf{R}$ wherein all of the parameters are unique. In certain applications, however, we may want to change our assumptions about the forms for $\mathbf{R}$. For example, we might have good reason to believe that all of the observations have different error variances and they were independent of one another (e.g., different methods were used for sampling), in which case

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 & 0 & 0 & 0 \\ 0 & r_2 & 0 & 0 & 0 \\ 0 & 0 & r_3 & 0 & 0 \\ 0 & 0 & 0 & r_4 & 0 \\ 0 & 0 & 0 & 0 & r_5 \end{bmatrix}.$$

Alternatively, we might have a situation where all of the observation errors had the same variance r, but they were not independent from one another. In that case we would have to include a covariance parameter k, such that

$$\mathbf{R} = \begin{bmatrix} r & k & k & k & k \\ k & r & k & k & k \\ k & k & r & k & k \\ k & k & k & r & k \\ k & k & k & k & r \end{bmatrix}.$$

R code

For this example, we make use of the *MARSS* package (Holmes et al. 2012) for *R* to estimate the parameters in the DFA model. To do so, we must specify the forms for the various vectors and matrices in a model list.

```r
library(MARSS)
## matrix dims
nn <- 5
mm <- 3
## model list
mod_list <- list(
  ## state eqn
  B = diag(mm),
  U = matrix(0,mm,1),
  Q = diag(mm),
  ## observation eqn
  Z = matrix(list(0),nn,mm), # all 0's for now
  A = matrix(0,nn,1),
  R = matrix(list(0),nn,nn)  # all 0's for now
)
## zero-out upper right corner of Z
mod_list$Z[,] <- paste0(rep(seq(nn),mm), rep(seq(mm), each=nn))
for(rr in 1:(mm-1)) {
  for(cc in (rr+1):mm) {
    mod_list$Z[rr,cc] <- 0
  }
}
## fill in diagonal of R
diag(mod_list$R) <- paste0(seq(nn),seq(nn))
#mod1 <- MARSS(rbind(x1, x2), model=mod_list, inits=list(x0=matrix(0,2,1)))
```