Theoretical study Quarto -document

2026-01-28

To do 22.2.

- Main consideration: Should measurement invariance be dropped from the common factor model and just call it autoregressive common factor model - or something similar. This would work as a simplification for the theoretical part.
  - Then in simulations, we turn to measurement invariance studies and introduce LMI. s-LMI being independent of stationarity makes a certain point on its usefullness. Maybe LMI models can in fact be used to study non-stationary processess with orthogonal factors, which can be then projected back (if necessary) ordinary covariance.
- Add/integrate previous Rmd file of theoretical result as extension to this one.
- Recheck delta t.
- Check if subindex VAR(1) is necessary for Sigma.
- Proofs, analysis…

Notes:

- Everything is a scratch and works as a basis for argument, which will be clarified when results are in. At this stage result is that drift matrix must be symmetric. This is already useful for simulations - and very interesting - but we'll see what other things turn out. I (Sakari) am not sure at this point still if VAR(1) can in facto produce covariance matrix compatible with a common factor model, though i consider it very likely.
- This file will be iterated and grows from autoregression -stuff onto the full article. Some foundational layouts are already put in place, but should not be treated as set.

Weak sense stationary VAR(1) and strict longitudinal measurement invariance

In the following VAR(1) process and strict longitudinal measurement invariance (s-LMI) are compared to each other by the covariance structure that VAR(1) imposes and how an s-LMI model fits to this (here equivalently 'is compatible' with) covariance. s-LMI makes sense as a theoretical model since it captures the simplest scenario of a true common factor model where only the common factor itself can change. VAR(1) is also a simple, if not the simplest, vector autoregression symptom-network model. CT-VAR and VAR can be linked through a transformation when fixed time intervals are used, or separate study for CT-VAR can be done.

Especially noteworthy results from mathematical analysis could be possible constraints that arise and how they can guide empirical simulations and tell us the reason why s-LMI might not be compatible with VAR(1) generated data in the first place. Alternatively, if some VAR(1) models can generate s-LMI compatible data, what constraints are necessary for the VAR(1) process to produce it?

Common Factor theoretical model A latent variable η is a causal entity which creates observable symptoms. Conditional independence of variables $P(X_i|\eta, X_j) = P(X_i|\eta) : i \neq j$ given the common factor η is the central proposition of the model. All changes in symptoms are due to changes in η. M_1 will represent an empirical (data-generating) model of the common factor theory.

Mutualistic symptom-network theoretical model Symptoms have causal effects on each others. Symptoms can change by themselves or change due to effects not within the symptom-network itself. The symptom-network is all the symptoms. M_2 will represent an empirical (data-generating) model of the symptom-network theory.

The models need to be kept vague for now, as the key parts of both differ in each scenario to be examined. Empirical distinguishability is here defined - for now - up to the fourth order moment:

Empirical distinguishability

$$\begin{aligned} E_{M_1}(X) - E_{M_2}(X) &= D_E \\ \mathrm{Cov}_{M_1}(X) - \mathrm{Cov}_{M_2}(X) &= D_C \\ \mathrm{Skew}_{M_1}(X) - \mathrm{Skew}_{M_2}(X) &= D_S \\ \mathrm{Kurt}_{M_1}(X) - \mathrm{Kurt}_{M_2}(X) &= D_S \end{aligned}$$

That is, after defining the models, data simulated from both of these models should not be the same w.r.t. their expectation, covariance, skewness and kurtosis. Our aim is to identify practical conditions, where the distinguishability is highest - i.e., when the differences in either expectation or covariance become largest. We do this in X different scenarios. In all scenarios we define a respective common factor theory model as well as a symptom-network

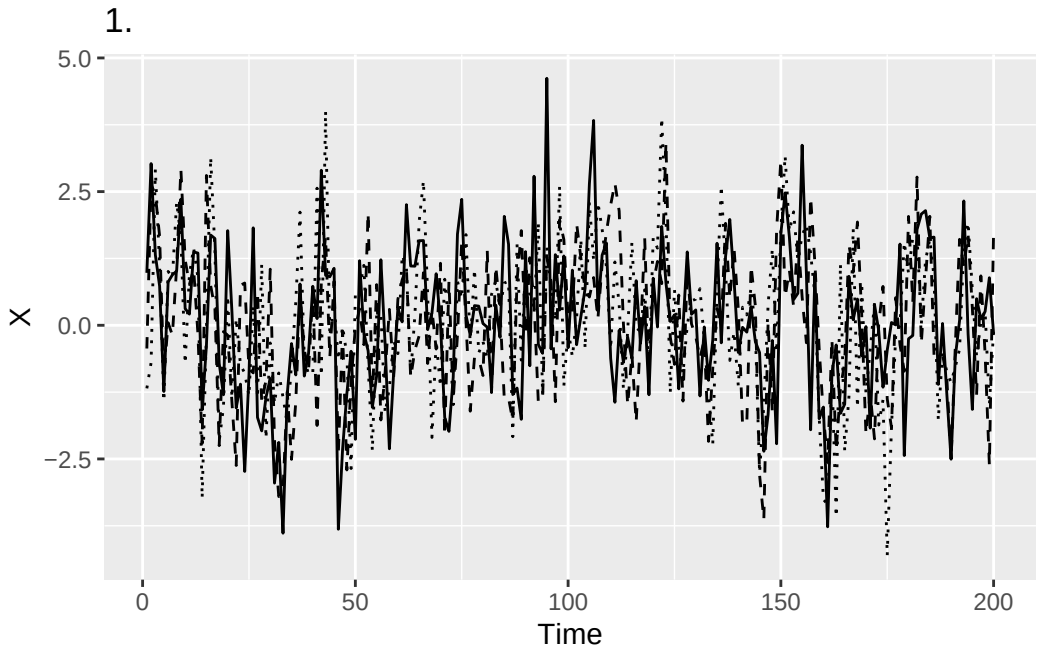
mutualism model. We then aim to analytically show the differences between the models, and then through simulation quantify what types of practical difference we might see. We'll be denoting symptoms as random matrix X, of which columns are the symptoms. We use two theoretical models:

The several scenarios relate to study designs, in which both models have been used using real world empirical data. First, we examine changes over time. Second, ..., X:th.

Change across time: Strict Longitudinal Measurement Invariance vs VAR(1) models

Regarding changes in symptoms across time, the most basic models that we can use for common factor and symptom-network theory are the strict longitudinal measurement invariance (s-LMI) referred to as and the vector autoregressive model of order one (VAR(1)). Here we will analytically approach differences between these models and then perform simulations to quantify how well empirically the models differ when generating from a mixture distribution. We will first inspect what type of covariance structure VAR(1) imposes, and then similarly s-LMI.

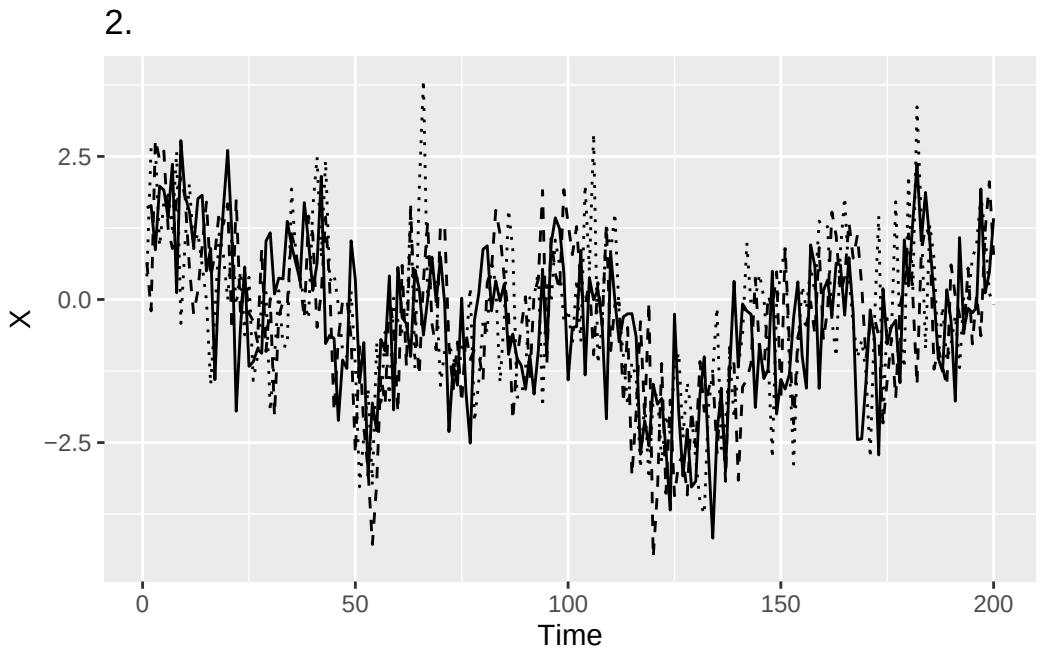
Their difference might not always be evident. See for example the figures 1. and 2. below, where data was generated from both models.



In Figure 1, the model was set up so that there is a true common factor η causing three variables X. The common factor had an autoregressive coefficient (of order 1) of $\sqrt{0.5} \approx 0.7071$ to itself and $\psi_t \sim N(0, \sqrt{0.5})$ innovations (0.5 variance). Factor loadings onto the observed variables

were 1. Each of the $X_{i,t}$ also had their own measurement errors $\omega_{i,t} \sim N(0, \sqrt{0.5})$. X_t has a covariance matrix approximately such that the diagonal elements are 2, and off-diagonal elements are 1.

If simulating from another numerical example, this time a stationary VAR(1) process, we can observe a fairly similar looking pattern.



The difference between the two example time series in D_C is

```
D_C = (cov(X_VAR) - cov(X_CF))
dimnames(D_C) = list(paste("X",1:3,sep = ""),paste("X",1:3,sep = ""))
round(D_C, 2)
```

```
      X1    X2    X3
X1 -0.23  0.06 -0.14
X2  0.06 -0.09 -0.07
X3 -0.14 -0.07 -0.12
```

The above is an ensemble covariance, that is, covariance calculated using all time points simultaneously. This is not the typical case however. More often we encounter cases where we observe few subsequent time points and then model them.

VAR(1) covariance structure

First VAR(1) imposed covariance is derived. Then s-LMI imposed covariance. Then we move on to inspect how they compare by equating them together to observe possible contradictions or restrictions. We begin with such a theoretical view on their imposed covariance structure, and later move to simulate data from both to see how distinguishable they are in commonly encountered longitudinal scenarios.

Before the theoretical analysis, a brief substantive consideration. We will assume (*as Jaakko suggested*) that the symptom-network represented as the VAR(1) model is a state. By this we mean that it is stationary. In addition to stationarity, we also assume that there are no external causes, which would make systematic changes to the symptoms and cause correlated innovations (or errors, innovations hereon). This assumption then gives us theoretical circumstance so that the innovations of the VAR(1) are independent of each other. While this assumption (that no external events affect the symptom-network) is unlikely to be true in a real-world scenario especially when considering longer periods of time (e.g., weeks, months), in any case theoretical analysis would become near impossible if no restrictions of this kind are made.

VAR(1) covariance structure at two subsequent time points

The VAR(1) model is defined in matrix format as $X_t = C + AX_{t-1} + \Gamma_t$, where Γ_t is independent error column vector with $E[\Gamma_t] = 0$, C is a constant assumed zero. Also assume centered X, $E[X_t] = 0$, in our case. Centering makes covariance calculations easier as the products of expected values can be mostly ignored (they become 0). A is $K \times K$ (borrowing from CT-VAR terminology) 'drift' matrix that includes all lagged effects of $K \times 1$ column vectors X_t to X_{t-1}, K being the number of observed items (symptoms). In this section the focus is on the $2K \times 2K$ covariance matrix where two subsequent measurement time points are observed. All matrices used are real-valued.

First, the covariance matrix (assumed stationary over time) is

$$\begin{aligned}
\mathrm{Cov}(X_t) &= E[X_t X_t^T] \\
&= E[(AX_{t-1} + \Gamma_t)(AX_{t-1} + \Gamma_t)^T] \\
&= E[AX_{t-1}X_{t-1}^T A^T] + \underbrace{E[\Gamma_t \Gamma_t^T]}_{=:\Psi} \\
&= AE[X_{t-1}X_{t-1}^T]A^T + \Psi \\
&= \Sigma_t =: \Sigma_{VAR(1)}
\end{aligned}$$

where stationarity poses that Σ is not dependent of time and so the covariance of VAR(1) is denoted as such from hereon. We will use $\Sigma_{VAR(1)}$ when we wish to be explicit about meaning the VAR(1) imposed within time point covariance. Γ_t is a random $K \times 1$ column vector of (serially) independent innovations at time point t, with $E[\Gamma] = 0, \mathrm{Var}(\Gamma = 1)$. Ψ is covariance

of the innovations within time point t - i.e., the contemporaneous covariance. We assumed that the innovations are independent, as described above, and so Ψ is diagonal.

The vectorized covariance matrix can be solved to equal

$$\text{vec}(\Sigma_{VAR(1)}) = (I - A \otimes A)^{-1}\text{vec}(\Psi)$$

Where vec is the vectorization operator and $\otimes$ is the Kronecker product. In the above the mixed Kronecker matrix vector product is used to obtain the result. (Derivation is currently in a scratch file.)

We will equate the VAR(1) posed $\Sigma_{VAR(1)}$ to s-LMI imposed covariance further down below.

Second, VAR(1) poses that observations at the time points X_t, X_{t-1} have covariance

$$\begin{aligned}\text{Cov}(X_t, X_{t-1}) &= E[X_t X_{t-1}^T] - E[X_t]E[X_{t-1}] \\ &= E[(AX_{t-1} + \Gamma_t)X_{t-1}^T] \\ &= E[AX_{t-1}X_{t-1}^T] + E[\Gamma_t X_{t-1}^T] \\ &= A\Sigma_{t-1} + E[\Gamma_t X_{t-1}^T]\end{aligned}$$

Independent errors means that $E[\Gamma_t X_{t-1}^T] = \text{Cov}(\Gamma_t, X_{t-1}) = 0$ leading to

$$\begin{aligned}\text{Cov}(X_t, X_{t-1}) &= A\Sigma_{t-1} \\ \text{Cov}(X_{t-1}, X_t) &= \Sigma_{t-1}^T A^T\end{aligned}$$

Where the two covariances above must be the same - i.e., the covariance matrix is symmetric. This means that every VAR(1) process implies that covariance of observations from two subsequent time points $t, t-1$ is

$$Cov((X_{t-1}, X_t), (X_{t-1}, X_t)) = \begin{pmatrix} \Sigma_{t-1} & A\Sigma_{t-1} \\ \Sigma_{t-1}^T A^T & \Sigma_t \end{pmatrix}$$

Where the above covariance matrix is the $2K \times 2K$ covariance matrix of the observed data from the two time points. In addition, stationarity directly implies $\Sigma_t = \Sigma_{t-1}$. (For now, notation with sub-index i will be kept for clarity as it is. s-LMI is not necessarily imply stationary, so confusion might be avoided.)

s-LMI covariance structure at two subsequent time points

s-LMI with 1 common factor decomposes Σ_{t-1} into following

$$\Sigma = \Lambda\Lambda^T + \Omega_{t-1}$$

where by definition of s-LMI Ω_{t-1} is diagonal and and Λ is a $K \times 1$ column vector of factor loadings constant over time. We also need the covariance of the common factor at both time

points. Let δ be the latent regression coefficient which links the common factor to itself at a previous time point such that $\eta_t = \delta\eta_{t-1} + \psi_t$, where ψ_t is independent random term ('innovation', 'error', 'disturbance') with $E[\psi_t] = 0$. Assuming standardized common factor such that $E[\eta_{t-1}] = 0$, $Var(\eta_{t-1}) = 1$ covariance of the common factor at two subsequent time points is

$$\begin{aligned}\text{Cov}(\eta_{t-1}, \eta_t) = E[\eta_{t-1}\eta_t] - E[\eta_{t-1}]E[\eta_t] = \\ E[\eta_{t-1}(\delta\eta_{t-1} + \psi_t)] = \\ E[\delta\eta_{t-1}^2 + \eta_{t-1}\psi_t] = \\ \delta Var(\eta_{t-1}) = \delta\end{aligned}$$

Now lets look at the $2K \times 2K$ covariance matrix from the perspective of strict LMI. A s-LMI model imposes that

$$\text{Cov}((X_t, X_{t-1}), (X_t, X_{t-1})) = \begin{pmatrix}\Lambda & 0 \\ 0 & \Lambda\end{pmatrix}\begin{pmatrix}1 & \delta \\ \delta & \delta + Var(\psi_t)\end{pmatrix}\begin{pmatrix}\Lambda^T & 0 \\ 0 & \Lambda^T\end{pmatrix} + \begin{pmatrix}\Omega_{t-1} & \Omega_{across} \\ \Omega_{across}^T & \Omega_t\end{pmatrix}$$

where

$$\begin{pmatrix}\Lambda & 0 \\ 0 & \Lambda\end{pmatrix}$$

is a block matrix that sandwiches the 2×2 covariance matrix of the common factor at both time points.

$$\begin{pmatrix}\Lambda & 0 \\ 0 & \Lambda\end{pmatrix}\begin{pmatrix}1 & \delta \\ \delta & \delta + Var(\psi_t)\end{pmatrix}\begin{pmatrix}\Lambda^T & 0 \\ 0 & \Lambda^T\end{pmatrix} + \begin{pmatrix}\Omega_{t-1} & \Omega_{across} \\ \Omega_{across}^T & \Omega_t\end{pmatrix} =$$
$$\begin{pmatrix}\Lambda & 0 \\ 0 & \Lambda\end{pmatrix}\begin{pmatrix}\Lambda^T & \delta\Lambda^T \\ \Lambda^T\delta & (\delta + Var(\psi_t)\Lambda^T\end{pmatrix} + \begin{pmatrix}\Omega_{t-1} & \Omega_{across} \\ \Omega_{across}^T & \Omega_t\end{pmatrix} =$$
$$\begin{pmatrix}\Lambda\Lambda^T + \Omega_{t-1} & \Lambda\Lambda^T\delta + \Omega_{across} \\ \Lambda\Lambda^T\delta + \Omega_{across}^T & \Lambda\Lambda^T(\delta + Var(\psi_t)) + \Omega_t\end{pmatrix}$$

From the above we see that the strict LMI can only be compatible with any process with stationary covariance, if $\delta + Var(\psi_t) = 1 \Rightarrow 1 - \delta = Var(\psi_t)$ (assuming Λ is non-zero). (When fitting a s-LMI model this is allowed.) We also see that s-LMI is compatible with non-stationary processes where the covariance is proportional to $\delta + Var(\psi_t)$ aligning with previous theoretical analysis where covariance increased over time in a LMI preserving model.

A brief note on notation: We'll be using simply Ω for the s-LMI residual covariance, since residual covariance is assumed invariant over time $\Omega_{t-1} = \Omega_t = \Omega_{t+1} = ... = \Omega$.

Using the above auxiliary results we can move to analyse the null hypothesis (hypotheses) of no difference between VAR(1) and s-LMI.

Working null hypothesis (1) If covariance at some (measurement) time point is perfectly explained by a common factor model with non-zero factor loadings Λ, and if the data generation is from a stationary VAR(1) model, then s-LMI model fits perfectly.

Considering only the subset of VAR(1) processes which create a covariance matrix that can be perfectly explained by a common factor model is done because we're interested in how (if at all) VAR(1) can deviate from s-LMI in terms of produced data. Understandably, if any VAR(1) model creates covariance structure incompatible with s-LMI model at some time point (i.e., a covariance matrix non-compatible with a common factor model), then deviation must occur (although the extent to which this occurs is not clear at this point).

If the VAR(1) generated $2K \times 2K$ matrix cannot be explained by the strict LMI model, this seems likely to be because the off diagonal blocks of covariance matrices across time points are non-compatible with the respective s-LMI model imposed across time covariance. Combined with the restriction on the within time point covariance, this might lead to contradictions.

This gives us the following null hypothesis (1) equations (from the $2K \times 2K$ matrices imposed by VAR(1) and s-LMI)

$$\begin{aligned}
(I - A \otimes A)^{-1}\text{vec}(\Psi) &= \text{vec}(\Lambda\Lambda^T + \Omega) \qquad \Rightarrow \\
(I - A \otimes A)^{-1}\text{vec}(\Psi) &= \text{vec}(\Lambda\Lambda^T) + \text{vec}(\Omega) \\
AE[X_{t-1}X_{t-1}^T]A^T + \Psi &= \Lambda\Lambda^T + \Omega \Rightarrow \\
A\Sigma A^T + \Psi &= \Sigma \Rightarrow \\
(A \otimes A)\text{vec}(\Sigma) &= \text{vec}(\Sigma) - \text{vec}(\Psi)
\end{aligned}$$

and

$$A\Sigma_{t-1} = \Lambda\Lambda^T\delta + \Omega_{across}$$

both of which must be true for the null hypothesis (1) to hold. Assuming that the null hypothesis (1) is true, further analysis of the respective equations show

$$\begin{aligned}
A\Sigma_{t-1} &= \Lambda\Lambda^T\delta + \Omega_{cross} \Leftrightarrow \\
A &= \Lambda\Lambda^T\delta\Sigma_{t-1}^{-1} + \Omega_{cross}\Sigma_{t-1}^{-1} \Leftrightarrow \\
A + \delta\Omega\Sigma_{t-1}^{-1} &= \Lambda\Lambda^T\delta\Sigma_{t-1}^{-1} + \delta\Omega\Sigma_{t-1}^{-1} + \Omega_{cross}\Sigma_{t-1}^{-1} \Leftrightarrow \\
A + \delta\Omega\Sigma_{t-1}^{-1} &= \delta \underbrace{(\Lambda\Lambda^T + \Omega)}_{=\Sigma_{t-1} \text{ by assumption}} \Sigma_{t-1}^{-1} + \Omega_{cross}\Sigma_{t-1}^{-1} \Leftrightarrow \\
A &= \delta I + (\Omega_{cross} - \delta\Omega)\Sigma_{t-1}^{-1}.
\end{aligned}$$

We'll see if contradictions arise when including three time points in the analysis below. But first, we need to do some generalizations and discuss the implications of VAR(1) model and the null hypothesis (1) further.

VAR(1) covariance compared to s-LMI covariance at T time points.

As the scenario where two subsequent measurement points are observed is possibly the most common one, we'll keep the above discussion in place for now. On the other hand, it is of interest to analyze what happens when multiple subsequent time points are included. This

is perhaps less common in measurement invariance literature, but more common in VAR literature.

We have (proven below) that as the distance in time between two time points $\Delta t = (t_i - t_j) \to \infty, i > j$ samples from VAR(1) model generated data at those two time points have 0 covariance. This would make the asymptotic $2K \times 2K$ perfectly explained by a measurement invariance model, because the main diagonal covariance matrices $\Sigma_{t-1} = \Sigma_t$ are perfectly explained by a common factor model by assumption, and the off diagonal matrices are 0, which is allowed in a strict LMI model (no cross-covariance between the observed time points and 0 regression coefficient for the latent variable). Again, we are assuming that VAR(1) model generated Σ perfectly compatible with s-LMI. This is a less practically meaningful case arguably since observed data with 0 across time point covariance is not common. This also does prove that a VAR(1) process could in some sense be the true data generating process even if no across time point covariance is observed (no lagged effects are estimated) if one were to claim that Δt is very small: We're just not observing time points close enough to each other to see the VAR(1).

Nevertheless - considering that the above scenario is not a typical one in psychopathology research - we can attempt to generalize the above results concerning the $2K \times 2K$ matrix to an $TK \times TK$ matrix where we have T measurement of K symptoms over occasions at constant time intervals. Brief notes are made as we move on and a summary at the end.

Here we change the notation a little and use an arbitrary time point t as the 'first' measurement time point, so that time points increase $t, t+1, t+2, ..., T$, where T is the last measurement time point. The $TK \times TK$ matrix is

$$\begin{pmatrix} \Sigma_t & \dots & \Sigma_{t,T} \\ \vdots & \ddots & \vdots \\ \Sigma_{t,T}^T & \dots & \Sigma_T \end{pmatrix}$$

Proceeding again from VAR(1) to s-LMI and then equating between the models. Let $^{(T)}$ denote the matrix raised to power of T. The $\Sigma_{t,T}$ for VAR(1) is

$$\Sigma_{t,T} = \text{Cov}(X_t, X_T) = A^{(T)}\Sigma_{VAR(1)}$$

Proof of the asymptotic 0 across time point covariance. On a brief note we can further decompose the above equation, using the power method of eigenvalues, into

$$PD^{(T)}P^{-1}\Sigma_{VAR(1)}$$

where D is a diagonal matrix of eigenvalues of A, P is and orthonormal matrix of eigenvectors of A as columns. From this decomposition we directly see the above mentioned asymptotic property that as distance in time between time points increases, D is raised to a larger power

and decreases eventually to the zero matrix. This is because eigenvalues of (stationary) A are less than one, meaning that all diagonal elements of $\text{diag}(D) : |d_{ii}| < 1$ as well, and hence the matrix power converges to 0. This means that a true data generating stationary VAR(1) model should produce across time point covariances that reduce to zero as distance in time between any two time points increases (practically speaking perhaps as years pass). Such observations might be sparse in psychopathology literature and it is accepted in the literature that even if any autoregressive model would be the true data generating model, it would unlikely be stationary over lengthier time periods (e.g., years again). This property of stationary VAR(1) does - as we have – restrain the analysis to a psychopathological state, but this contemplation is omitted here.

For s-LMI, respectively, we'll use $\delta_2, \delta_3, ..., \delta_t, \delta_{t+1}, ..., \delta_T$ to denote the regression coefficient between subsequent time points (there is no regression at the first time point in s-LMI). Also, the $\Omega_{t,t+1}$ needs to be generalized to include residual covariances between any two time points so that $\Omega_{t,T}$ is the residual covariance between time point t and time point T. Ω is the within time point residual covariance invariant over time.

Few more generalizations and constraints before we can equate the VAR(1) implied covariance to the s-LMI implied covariance again using multiple time points. We have that the VAR(1) imposed covariance between any two subsequent time points is the same. This also generalizes to the VAR(1) imposed covariance between any two equidistant time points as they are $A^{\Delta t}\Sigma$, which only depends on the distance in time. This means that δ_t must be a constant since otherwise the s-LMI imposed covariance between two subsequent time points $\Lambda\Lambda^T\delta + \Omega_{t,t+1}$ would not be the same. More precisely, $\Omega_{t,t+1}$ is by definition diagonal and so can only change the diagonal to some extent - off-diagonal elements would not be the same. This also means that $\Omega_{t,t+1}$ is the same for any t.

Using that δ is constant (and other assumptions established above), the covariance between the common factor to itself between any time points two time intervals apart from each other is

$$\begin{aligned}
\text{Cov}(\eta_t, \eta_{t+2}) &= E[\eta_t \eta_{t+2}] \\
&= E[\eta_t(\delta_{t+2}\eta_{t+1} + \psi_{t+2})] \\
&= E[\eta_t(\delta_{t+2}(\delta_{t+1}\eta_t + \psi_{t+1}) + \psi_{t+2})] \\
&= E[\eta_t(\delta_{t+2}\delta_{t+1}\eta_t + \delta_{t+2}\psi_{t+1} + \psi_{t+2})] \\
&= E[\delta_{t+2}\delta_{t+1}\eta_t\eta_t] + E[\delta_{t+2}\psi_{t+1}\eta_t] + E[\psi_{t+2}\eta_t] \\
&= \delta_{t+2}\delta_{t+1} = \delta^2
\end{aligned}$$

and for three time intervals apart

$$
\begin{aligned}
\text{Cov}(\eta_t, \eta_{t+3}) &= E[\eta_t \eta_{t+3}] \\
&= E[\eta_t(\delta\eta_{t+2} + \psi_{t+3})] \\
&= E[\eta_t(\delta(\delta\eta_{t+1} + \psi_{t+2}) + \psi_{t+3})] \\
&= E[\eta_t(\delta(\delta(\delta\eta_t + \psi_t) + \psi_{t+2}) + \psi_{t+3})] \\
&= E[\eta_t(\delta(\delta^2\eta_t + \delta\psi_t) + \psi_{t+2}) + \psi_{t+3})] \\
&= E[\eta_t(\delta^3\eta_t + \delta^2\psi_t + \delta\psi_{t+2} + \psi_{t+3})] \\
&= E[\eta_t\delta^3\eta_t + \eta_t\delta^2\psi_t + \eta_t\delta\psi_{t+2} + \eta_t\psi_{t+3})] \\
&= \delta^3 = \text{Cov}(\eta_t, \eta_{t+2})\delta
\end{aligned}
$$

which can be by induction (not currently properly done) shown to result in

$$
\text{Cov}(\eta_t, \eta_{t+\Delta t}) = \delta^{\Delta t}
$$

This implies that the covariance between any two time points must be $\Lambda\Lambda^T\delta^{\Delta t} + \Omega_{t,T}$.

From the previous result for the $2K \times 2K$ matrix we can then generalize to the $TK \times TK$ matrix

$$
\begin{pmatrix}
\Lambda\Lambda^T + \Omega & \dots & \Lambda\Lambda^T\delta^{(T-1)} + \Omega_{1,T} \\
\vdots & \ddots & \vdots \\
\Lambda\Lambda^T\delta^{(T-1)} + \Omega_{1,T}^T & \dots & \Lambda\Lambda^T(\delta + Var(\psi_t)) + \Omega
\end{pmatrix}
$$

Now we can obtain the more general, null hypothesis equation relating the across time point covariances and within time point covariances for any lag

$$
\begin{aligned}
\Sigma_{\Delta t} &= \\
A^{(\Delta t)}\Sigma &= \Lambda\Lambda^T\delta^{\Delta t} + \Omega_{\Delta t}
\end{aligned}
$$

This directly tells us that A must be symmetric, positive definite:

$$
\begin{aligned}
A^{(\Delta t)}\Sigma &= \Lambda\Lambda^T\delta^{\Delta t} + \Omega_{\Delta t} \\
\Rightarrow A^{(\Delta t)} &= (\Lambda\Lambda^T\delta^{\Delta t} + \Omega_{\Delta t})\Sigma^{-1}
\end{aligned}
$$

where the right hand side is a multiplication of two positive-definite (covariance) symmetric square matrices, which must also be positive definite and symmetric.

Discussion notes

The above result has the substantive implications, that a symptom-network VAR(1) model can only be compatible with a common factor model in the restricted case of all symptoms affecting each other with bi-directional effects which are equivalently strong in each direction $X_i \rightarrow X_j = X_i \leftarrow X_j$. That is, distinguishability D_C will be larger the more asymmetric the drift matrix A is.

Looking from a measurement perspective, we might be interested what happens in a 0 measurement error scenario. So assume there is no unique variance in X. In this case we obtain

$$\begin{aligned} A^{(\Delta t)}\Sigma &= \Lambda\Lambda^T\delta^{\Delta t} + \underbrace{\Omega_{\Delta t}}_{=0} \\ \Rightarrow A^{(\Delta t)}\Sigma &= \underbrace{\Lambda\Lambda^T}_{\text{By assumption: } \Sigma-\Omega=\Sigma}\delta^{\Delta t} \\ \Rightarrow A^{(\Delta t)}\Sigma &= \delta^{\Delta t}\Sigma \\ \Rightarrow A^{(\Delta t)} &= \delta^{\Delta t} I \end{aligned}$$

that is, A must be a scalar multiple of the identity matrix. This would lead to a contradiction (if we assume that there must be positive covariance between symptoms at least), since now there should be 0 covariances between symptoms.

Without measurement error

WE SHOULD ASSUME NO SERIAL COVARIANCE OF ERRORS! If we simplify, and assume that there is no unique variance ($\Omega = 0$), then

$$\begin{aligned} A^{(\Delta t)}\Sigma &= \Lambda\Lambda^T\delta^{\Delta t} + \underbrace{\Omega_{\Delta t}}_{=0} \\ \Rightarrow A^{(\Delta t)}\Sigma &= \underbrace{\Lambda\Lambda^T}_{\text{By assumption: } \Sigma-\Omega=\Sigma}\delta^{\Delta t} \\ \Rightarrow A^{(\Delta t)}\Sigma &= \delta^{\Delta t}\Sigma \end{aligned}$$

We see that A must hence be scalar multiple of the identity matrix (or a scalar). This is a contradiction, because a diagonal A and diagonal Ψ make covariances between the variables zero. This means that Λ must also be zero. This is against our assumption of null hypothesis (1).

If measurement error exists ($\Omega \neq 0$)

$$A^{(\Delta t)}\Sigma = \Lambda\Lambda^T\delta^{\Delta t}$$

where we observe that the left hand side is rank 1, since $\Lambda\Lambda^T$ is a product of two column vectors. $\Lambda\Lambda^T$ has one non-zero eigenvalue (the magnitude of the common factor).

Rewriting and differentiating w.r.t. Σ we obtain that

$$
\begin{aligned}
& A^{(\Delta t)}\Sigma = \Lambda\Lambda^T\delta^{\Delta t} + \Omega_{\Delta t} \\
\Rightarrow & A^{(\Delta t)}\Sigma = (\Sigma - \Omega)\delta^{\Delta t} + \Omega_{\Delta t} \\
\Rightarrow & A^{(\Delta t)}\Sigma = (\Sigma - \Omega)\delta^{\Delta t} + \Omega_{\Delta t} \\
\Rightarrow & A^{(\Delta t)}\Sigma = \Sigma\delta^{\Delta t} - \Omega\delta^{\Delta t} + \Omega_{\Delta t} \quad (\frac{\partial}{\partial\Sigma}) \\
\Rightarrow & A^{(\Delta t)} = \delta^{\Delta t}
\end{aligned}
$$

We observe that A must be diagonal, and this leads again to a contradiction.

Summary and implications OLD 31.1.2024

- A direct observation from the null hypothesis (1) and auxiliary null hypothesis (A1) equations shown above is that A must always be symmetric for the common factor s-LMI model to be compatible with stationary VAR(1) generated data. Hence, asymmetry in A produces incompatibility to s-LMI.
- s-LMI is compatible with the idea that VAR(1) poses that across time point covariance approaches 0 as the distance between two time points increases. This scenario might not be frequently observed however, suggesting that - at least some part of - psychopathology processes cannot be understood as vector autoregressive processes. This is true for any stationary process to my knowledge, as the requirement is that the process will necessarily remain bounded to some extent to its 'state'. In fact, it might be reasonable to integrate both viewpoints as is done in trait-state models.

Next

- Perhaps one possibility is something like Chi-squared testing with vectorized (non-redundant) squared elements of $A^T_{lower-tri} - A_{upper-tri}$.
- At this point it still is not evident that VAR(1) can generate a covariance matrix perfectly decomposable to a common factor model(?). Even more, if this is possible for multiple subsequent timepoints is unclear.
- Also it is not clear what happens when sampling from a population at different time points so that subjects are sampled at different time lags.
- Stationarity is also not necessarily a condition which should be imposed, which can be discussed further.
- VAR(1) is a special case of CT-VAR. The respective transformations from CT-VAR to VAR(1) are available.

- Measures of asymmetry such as s

$$s \equiv (|A_{sym}| - |A_{anti}|)/(|A_{sym}| + |A_{anti}|) \tag{1}$$

  where A is decomposed into its symmetric and asymmetric parts can be used also. The metric above is shown at: stack exchange. Asymmetry of A could further be approached analytically for example by decomposing A into its symmetric and asymmetric parts, and/or through simulations where asymmetry of A is varied by producing random matrices with some logic with how asymmetric is produced in A.
-