

Statusupdate: Simulationsstudie

Table of contents

1 Einleitung	1
Simulation Parameters	1
Modellgleichungen (Kurzform)	2
2 Probleme mit positiven Verteilungen	2

1 | Einleitung

Ziel ist es, zu prüfen, wie stark misspezifizierte Residualverteilungen (normal) die Schätzung von VAR/DSEM Koeffizienten in VAR/DSEM verzerrt, wenn die wahre Verteilung non-normal ist (sowohl in den Randverteilungen wie auch der Copula). Untersuchen wollte ich, wie besprochen, 2 Modelltypen:

1. DSEM – random intercepts (sprich, personenspezifisch) und random slopes, also AR und cross lagged effects.
2. Einfaches VAR(1) – gleiche Gleichung, aber ohne random effects.

Beide Modelle werden jeweils mit (a) klassischer Normalverteilungsannahme (multivariat normal -> misspecified) und (b) dem korrekten Modell -> Copula-Variante (Gaussian oder Clayton) mit entsprechend korrekt spezifizierten Residualverteilungen.

Simulation Parameters

Faktor	Ausprägungen
Schiefe-Parameter α	-9 (linksschief) / +9 (rechtsschief)
Personen N	50 / 300
Zeitpunkte T	30 / 100
Diagonale AR	0.2 / 0.5
Cross-Lag	0.0 / 0.3

Faktor	Ausprägungen
Kendall- τ (Copula)	0.2 / 0.5
Replikationen	X

- Die Residualvarianz wird pro Bedingung so skaliert, dass die angestrebte effektive Autokorrelation erhalten bleibt
- im Augenblick gibt es einen burn in von 30

Modellgleichungen (Kurzform)

Ebene	Gleichung
Level 1	$y_{it}^{(W)} = \Phi_i y_{i,t-1}^{(W)} + \varepsilon_{it}$
Level 2	$\mu_i \sim N(\gamma_\mu, \Sigma_\mu) \text{vec}(\Phi_i) \sim N(\gamma_\Phi, \Sigma_\Phi)$

Bzw. nur level 1 für VAR Modell.

2 | Probleme mit positiven Verteilungen

Für die Modellierung der Residuen, die nur positive Werte annehmen, sollten Verteilungen wie die Exponential- oder Gammaverteilung verwendet werden. Exponential-, Gammaverteilungen kriege ich nicht implementiert: Da schlicht numerisch instabil.

Die Posteriorschätzung macht das einfach nicht mit, da der untere Grenzwert einen harten cut in der Verteilung verursacht. Das DSEM ist schon komplex und zusammen mit der Copula führt das zu hunderten von Divergenzen pro chain (selbst mit adapt_delta auf 0.99). Ich habe auch verschiedene Reparametrisierungen versucht, beispielsweise die Werte zu transformieren oder nicht-zentrierte Ansätze für die hierarchischen Parameter zu verwenden. Hat das Problem aber nicht behoben. Entweder blieben die Divergenzen bestehen oder die ESS-Werte waren niedrig, sobald es anspruchsvoller wurde.

Ich bin dann erst einmal einen Schritt zurückgegangen und habe mich auf die VAR-Modelle konzentriert. Für die VAR-Modelle untersuche ich jetzt die Performance mit schiefverteilten Residuen, die einer Skew-Normal-Verteilung folgen, für die einzelnen Zeitreihen. Hier gibt es auch Probleme, die ich dann gerne besprechen würde.

Im Augenblick schau ich mir also VAR Modelle an mit gaussian/clayton copula und skew normal residuen.