

Introduction to Bayesian structural equation modelling

Will Hardy

2020-05-23

Contents

1	Prerequisites	5
2	Introduction to structural equation modelling	7
2.1	Specification	7
2.2	Identification	8
2.3	Estimation	8
2.4	Data collection and preparation	9
2.5	Respecification	9
2.6	Reporting results	9
3	Introduction	11
3.1	Priors	12
3.2	Estimating parameters	13
3.3	Regression exercise	15
3.4	References	15
4	Introduction	17
4.1	What does that mean?	17
4.2	Model fit	18
4.3	References	19
5	Running BSEM analyses in AMOS	21

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

Chapter 2

Introduction to structural equation modelling

- What is structural equation modelling (SEM)?
 - SEM is a variety of methods that analyse covariance matrices¹.
- combination of factor analysis and path modelling
- There are six basic steps for most SEM analyses:
 1. Specify the model.
 2. Evaluate model identification (if not identified then return to step 1).
 3. Select the measures (operationalise the constructs) and collect, prepare, and screen the data.
 4. Estimate the model.
 1. Evaluate model fit; if it is poor then respecify the model if doing so is justifiable (skip to step 5); otherwise retain no model (skip to step 6).
 5. Respecify the model, which is assumed to be identified (return to step 4).
 6. Report the results.

2.1 Specification

- Hypotheses are often expressed with graphical conceptual models. These provide a visual representation of theoretical variables of interest and expected relationships between them.
- Outcome variables in SEM are referred to as *endogenous* variables.

¹Covariance is...

- All must have at least one cause in the model.
- Some of these causes are independent variables.
- Independent variables in SEM are known as *exogenous* variables.
 - The causes of these are not represented in the model.
- Specification is the most important step.
 - Results from subsequent steps assume that the model/the hypotheses are basically correct.
- Making a list of changes, that are justifiable according to theory or empirical results, at this stage can be useful for step five.

2.2 Identification

- Graphical conceptual models are useful but need to be translated into statistical models that can be analysed.
- Statistical models must respect certain rules/restrictions. One of these is identification.
- A model is identified if it is *theoretically* possible for the computer to derive a unique estimate of every model parameter. This is a property of the model, not of the data.
 - A researcher may specify a model that is true to a particular theory, but is not identified. To analyse such a model in SEM requires it to be respecified, but this may result in a model that is akin to making an intentional specification error in relation to the theory.

2.3 Estimation

- A computer SEM too will be used to conduct the analysis (e.g., Mplus, lavaan in R, AMOS).
- Evaluate model fit.
 1. How well does the model fit the data?
 - When a model is a poor fit to the data, consider if respecification of the model can be justified given relevant theory and results of prior empirical studies.
 2. If/when a well fitting model is found, then interpret the parameter estimates.
 3. Consider equivalent or near equivalent models.

- An equivalent or near equivalent model may explain the data just as well as the researcher's preferred model, but with a contradictory pattern of causal effects among some of the variables.
- Researchers should explain why their preferred model should not be rejected in favour of equivalent ones.

2.4 Data collection and preparation

2.5 Respecification

- Researchers arrive at this step when initial model fit is poor.
- The list of justifiable modifications created earlier will be useful at this stage.
- Modifications should be guided by rational rather than statistical considerations.

2.6 Reporting results

- Many papers do not report results properly.
- There are guidelines available (??, p 464).

Chapter 3

Introduction

- What is it?
 - Bayesian statistics assumes that all parameters (e.g., correlation coefficients) have a distribution.
 - Frequentist statistics (a.k.a. classical or “normal” statistics) assumes that there is one true population parameter (e.g., there is one true correlation coefficient).
 - $\text{Posterior} = (\text{prior} * \text{likelihood}) / (\text{probability distribution})$.
 - The probability distribution is only important with multiple hypotheses or datasets, so in many instances the formula can be simplified to: $\text{posterior} = \text{prior} * \text{likelihood}$ or even more simply, $\text{posterior} = \text{what do we know before hand} * \text{what does the data say}$.
- Why would I use it?
 - Why not? Just because it is not “normal” is not a reason to use it as long as you effectively communicate your results.
 - Bayesian statistics allows you to:
 - * Incorporate (un)certainty in your model.
 - * Update your knowledge.
 - Some of the statistics are easier to interpret.
 - * There are no p-values.
 - * Credibility intervals (Bayesian) vs confidence intervals (Frequentist).
 - A 95% confidence interval: in the long run, 95% of the confidence intervals will contain the true population value.
 - A 95% credibility interval: there is 95% probability that the population value lies within the upper and lower bounds.
 - Technical reasons (e.g., identifying complex models, improving convergence, more accurate parameter estimates, non-normal data, missing data handling).

- Small samples.
 - * Bayes is not based on central limit theorem.
 - * It is also not a magical solution to small sample problems.

3.1 Priors

- Specific to populations - different populations will need different priors.
- Information about the parameter.
- Usually made up of three components:
 - Shape (e.g., normal).
 - Location (e.g., mean).
 - Certainty (e.g., standard deviation).
- Take the scale of the data, so if you want to use standardised effects then you must use standardised data.
- There are lots of different names for types of priors, however they can be split into three main types:
 - Uninformative/flat prior.
 - Weakly informative.
 - Strongly informative.
- Effects on the posterior:
 - Priors need be specified before looking at the data as they will influence the results of any analyses.
 - Strongly informative priors will result in a *prior driven posterior* where the results will be closer to the prior than the likelihood.
 - Weakly informative priors will result in a *data driven posterior* where the results will be close to the likelihood than the prior.
 - Sample size is also important, the larger the sample, the larger the influence of the likelihood on the posterior.
 - <https://www.rensvandeschoot.com/fbi-the-app/>
- Where do priors come from?
 - Prior research (e.g., meta analyses, individual studies, pilot studies).
 - Expert elicitation.
 - Software defaults if you don't specify them yourself.
 - * Different software = different defaults (van Erp et al., 2018).
- Prior data conflict.
 - Can be because the prior chosen does not represent the sample populations (e.g., the prior is based on the general population and your sample population is the “healthy ageing population”).
 - Can be of interest and the outcome of a study (e.g., the ability of teachers to predict students performance).

- Only likely to happen when you use strongly informative priors.
- Documenting priors.
 - You should explain which priors were chosen and why.
 - This should be documented and provided with any analyses.

3.2 Estimating parameters

- Gibbs sampler.
 - Estimates the parameter assuming that the other parameters are known.
 - * Using a random start value, a parameter can be estimated. This parameter estimate is then passed to the next iteration of computation to replace the random start value.
 - Simplifies computation.
 - Chains represent starting values.
 - Using multiple chains reduces the likelihood of accepting *local convergence*.
 - * A single chain may exhibit a stable mean and variance for a sufficient number of iterations to be accepted given a set of convergence criteria. However, that chain may, given enough time, change its mean. Two chains with different start values are less likely to display local convergence than one chain is, the probability of three chains displaying local convergence is smaller again, and so it goes on.
 - *Burn-in* phase to remove the influence of starting values.
 - More chains and/or more iterations allow for greater precision of estimates.
 - Want stable means and variances within trace plots.
 - If the model is correctly specified, it will *eventually* identify the target distribution.
 - Convergence.
 - * Gibbs sampler must run for t iterations for the burn-in phase, where t is great enough to remove the influence of the start values.
 - * Convergence is when all chains reach the same target distribution (i.e., there would be no way to tell which chain the sample draw was from).
 - * To check that convergence is stable, it is normal to double the number of iterations to check that convergence remains.
 - * 9/10 times increasing the number of iterations will solve convergence issues.

- * Including trace plots with manuscripts will allow others to (visually) assess convergence.
- * It is also important to inspect posterior distributions - Do they look like you expected them to? Do they make sense?
- * Autocorrelation is when an iteration is dependent on the last. One way to deal with auto correlation is through *thinning* where you only retain the n th sample. However, this does not acutally change the estimate and some would argue that it is unnecessary if the trace plots exhibit stable means and variances.
 - Thinning doesn't get rid of autocorrelation, it hides it. If using thinning, you require more iterations to recover precision.
- There are several ways to assess convergence (Kaplan and Depaoli, 2012):
 - * Inspecting trace plots for stability and good mixing.
 - * Potential Scale Reduction Factor (PSR; Gelman and Rubin, 1992). Different cut-off criteria, usually >1.1 or >1.05 .
 - In Mplus, with a single chain, PSR is defined using the third and fourth quarters of the chain.
 - * Kolmogorov-Smirnov (K-S) tests: a test of equality of the posterior parameter distributions across the different chains using draws from the chains.
 - * Geweke z-statistic: compares the first and last half of the post burn-in iterations. Significant test is evidence of local convergence.
- Sensitivity to priors.
 - The effect of priors on the posterior distribution was described above. It is important to provided evidence of how robust the estimates are, or how *sensitive* they are to the prior specification.
 - **This is not an attempt to find the “best” prior (e.g., the one that provides a posterior that supports your hypotheses).** And that is why it is important to specify your priors **before** you conduct any analyses.
 - Analyses should be re-run with both more and less informative priors. Differences in parameter estimates should be reported both relative to the original estimate and in absolute values. This may be a sentence or short paragrah in the results, supported with supplementary information.
- Depaoli and van de Schoot (Depaoli and van de Schoot, 2017) provide a checklist to help researchers avoid the naïve use of Bayesian statistics.

3.3 Regression exercise

<https://www.rensvandeschoot.com/tutorials/> > WAMBS checklist > exercise
in programme of your choice.

3.4 References

Chapter 4

Introduction

- What is Bayesian structural equation modelling (BSEM)? In their paper introducing BSEM, Muthén and Asparouhov (2012, p 316) stated that:

It is intended to produce an analysis that better reflects the researcher's theories and prior beliefs. It does so by systematically using informative priors for parameters that should not be freely estimated according to the researcher's theories and prior beliefs. In a frequentist analysis, such parameters are typically fixed at zero or are constrained to be equal to other parameters. In key applications, freeing these parameters would, in fact, produce a nonidentified model. The Bayesian analysis, however, identifies the model by substantively driven small-variance priors. It should be recognized that BSEM refers to the specific Bayesian approach proposed here of using informative, small-variance priors to reflect the researcher's theories and prior beliefs. Typically, this would be combined with the use of noninformative priors for parameters that would not be restricted in a corresponding ML analysis. For example, major loadings would have a normal prior with a very large variance. The BSEM approach of using informative priors is applicable to any constrained parameter in an SEM.

4.1 What does that mean?

- Typical maximum-likelihood (ML) approaches to confirmatory factor analysis (CFA) only allow indicators to load on their intended factors and fix cross-loadings (i.e., loadings onto non-intended factors) and the correlation between residual-error terms to zero. This approach is said to

be overly restrictive and often leads to rejection of models based on the Chi-square test of good fit (Marsh et al., 2009).

- Researchers therefore tend to rely on other fit-indices to justify accepting a model when the Chi-square test suggests that they should not, as they suggest it is overly sensitive to trivial misfit (Fong and Ho, 2013). However, these indices will also often reject the model, especially when there are a large number of indicators (Marsh et al., 2004). This can lead to researchers relaxing conventional criteria for model fit and/or modifying their model post-hoc to improve the fit, which can lead to the acceptance of a poor model or one with a fit that has been improved by chance based on iterative model respecification using modification indices (MacCallum et al., 1992).
- Whilst cross-loadings and correlated errors terms can be specified in ML-CFA, specifying too many will often lead to an underidentified model (Asparouhov and Muthén, 2009).
- Exploratory structural equation modelling is one alternative to traditional CFA approaches that is less restrictive (i.e., it allows for cross-loadings), however it still does not allow for correlated residual error terms, nor does it allow it to specify how close to zero the cross-loadings should be (Muthén and Asparouhov, 2012).
- BSEM allows the specification of both cross-loadings and residual correlations that are *approximately zero* this is done by using small-variance (i.e., strongly informative) priors on these parameters. What approximately zero means is defined by the user and may be informed by theory, however if these are large variance priors the model may be underidentified (Muthén and Asparouhov, 2012).
- The user may or may not specify priors for the intended factor loadings, if they are not specified, the software default will be used.
- It is important to consider priors before analysing the data, especially in small samples, as there is growing evidence that the “naïve” use of priors can lead to biased estimates, sometimes worse than would be obtained through ML estimation (?).

4.2 Model fit

Model fit in BSEMs can be assessed in using a number of metrics:

- Number of free parameters
- Posterior Predictive p-value (PPp value)
- 95% Confidence Interval for the Difference Between the Observed and the Replicated Chi-Square Values
- Deviance Information Criterion (DIC)

4.3 References

Chapter 5

Running BSEM analyses in AMOS

1. Missing data
 1. Missing data can be dealt with prior to BSEM analysis in a variety of ways ranging from the simple (e.g., listwise deletion, mean imputation) to far more complicated methods (e.g., multiple imputation).
 2. It is beyond the scope of this course to cover the various techniques and when to use them.
 3. By default, AMOS will do XX when there is missing data.
2. Standardise variables
3. Model specification
 1. Major loadings only
 2. Major loadings + minor cross-loadings
 3. Major loadings + minor cross-loadings + approximate zero covariance
4. Additional output
 1. Indirect moments
5. Run
6. Prior specification
 1. Default for all variables in AMOS is an uninformative prior (uniform; -1^{38} to $+1^{38}$).
7. Convergence:
 1. As far as I can tell, AMOS only uses a single MCMC chain.
 2. PSR
 3. SC for individual parameters

4. Trace plots for each parameter
5. Posterior
 1. First and last box creates polygons for the first and last third of the samples
8. Fit:
 1. PPp
 2. Chi-square
 3. DIC
9. Parameter estimates
10. Sensitivity analyses
 1. More informative priors
 2. Less informative priors

Bibliography

- Asparouhov, T. and Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3):397–438.
- Depaoli, S. and van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, 22(2):240–261.
- Fong, T. C. T. and Ho, R. T. H. (2013). Factor analyses of the Hospital Anxiety and Depression Scale: A Bayesian structural equation modeling approach. *Quality of Life Research*, 22(10):2857–2863.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.
- Kaplan, D. and Depaoli, S. (2012). Bayesian structural equation modelling. In Hoyle, R., editor, *Handbook of Structural Equation Modelling*. Guilford Press, New York.
- MacCallum, R. C., Roznowski, M., and Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance.
- Marsh, H. W., Hau, K.-T., and Wen, Z. (2004). Structural Equation Modeling : A Modeling Latent Growth Curves With Incomplete Data Using Different Types of Structural Equation Modeling and Multilevel Software. *Structural Equation Modeling*, 11(3):452–483.
- Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., and Trautwein, U. (2009). *Exploratory structural equation modeling, integrating CFA and EFA: Application to students’ evaluations of university teaching*, volume 16.
- Muthén, B. and Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3):313–335.

- van Erp, S., Mulder, J., and Oberski, D. L. (2018). Prior sensitivity analysis in default bayesian structural equation modeling. *Psychological Methods*, 23(2):363–388.