Title: Seeing the Impossible: Visualizing Latent Variable Models in flexplavaan

Latent variable models (LVMs) are powerful and flexible tools that allow users to model sophisticated relationships involving variables that are not observed directly. Unfortunately, LVMs suffer from several serious limitations. Foremost among these limitations is that there is no intuitive way to visualize the data from LVMs because the variables of interest are unobserved. This makes it hard to encode information from the model, diagnose local and global misfit, and evaluate model assumptions. In this paper, we introduce flexplavaan, an R package designed to augment analyses in lavaan and blavaan. This package provides easy-to-use visuals that make it easy to encode LVMs, evaluate model assumptions, and diagnose misfit. In this paper, we develop the logic behind LVM visualizations and illustrate how flexplavaan is able to detect misfit and violated assumptions, as well as provide guidance on how best to modify models.

**Introduction**

Amidst the push for open science, some have called for greater use of visualization techniques (Fife, n.d.; Fife & Rodgers, 2019; Tay, Parrigon, Huang, & LeBreton, 2016)

* Visuals improve encoding (Correll, 2015)
* Visuals highlight misfit (Healy & Moody, 2014)
* Visuals are an essential component to evaluating model assumptions (Levine, 2018; Tay et al., 2016)

These are essential components of *any* modeling procedure

* Yet data visualization is more challenging when using analyses of unobserved variables (LVMs) that do not have unique scores that can be plotted
  + Note an important distinction between visualizing conceptual models (e.g., through path diagrams) and visualizing statistical models (e.g., through scatterplots)
  + Though the former are common, the latter are not (Hallgren, McCabe, King, & Atkins, 2019)
* LVMs rely on “latent” or unobserved variables (Bollen, 1989)
* How does one visualize something that cannot be observed?[[1]](#footnote-1)
* Yet visuals are especially critical for LVMs

**Why Visualizations are Critical for LVMs**

There are many longstanding issues associated with LVM that visualizations could help resolve

* 1. There is no good way to assess model fit in LVMs The best method for assessing the adequacy of model fit is a source of contention in the field (Barrett, 2007; Hayduk et al. 2007; Steiger 2007)
  + in regression models, there are relatively intuitive metrics for assessing global fit
    - (Adjusted) R2 is a global estimate of model effect size that behaves in expected ways (insensitive to sample size, number of variables (Harrell 2010)
    - F statistic also behaves as expected
      * large N = rejecting null model
  + SEM practitioners typically rely on global fit metrics, which have known weaknesses (Steiger 2007)
    - chi square is overly sensitive to sample size (Steiger 2007)
      * plausibly reject models that are good approximations to the “true” model (Steiger 2007)
      * fail to reject poor fitting models because of small sample sizes (Jiang & Yuan, 2017)
    - Model effect sizes also behave poorly
      * plethora of fit indices all sensitive to various features of a model (e.g., N, # of variables, number of paths) (West et al., 2012)
      * misfit in one area (e.g., structural components) can be masked by fit in another (e.g., measurement) (Kline 2016)
      * “rules of thumb” are routinely used, but probably shouldn’t be (Chen, Curran, & Bollen, 2008)
  + Visuals can provide concise representations of statistical models
    - users can determine at a glance how well the model fits
* 2. Diagnosing local misfit is also problematic (Thoemmes et al., 2018)
  + - not only problematic for that section of the model, but misfit in one part can spread throughout the rest of the model when using conventional maximum likelihood estimation procedures (Bollen, 2019)e.g., measurement misfit might inflate (or attenuate) structural components
  + various approaches to addressing local misfit
    - 1. modification indices (MacCallum et al., 1992)
      * MIs indicate how much the chi square would change if a constraint on a model is removed (e.g., by changing a path constrained to zero to be freely estimated)
      * Only works by recommending particular paths be freely estimated
        + will not work if significant structural changes are required (e.g., adding a new latent variable) (Hayduk, 1990)
      * data guided approach
        + can be sensitive to overfitting (MacCallum et al., 1992)
        + particularly when theoretical modifications are made
    - 2. studying residuals of correlation matrix (Bollen 1989; Kline 2016)
      * easy to detect misfit among *two* variables but not easy to see a pattern of misfit
        + requires a great deal of effort on the part of the researcher, depending on how complex the model is
  + Visuals will make both global and local misfit readily apparent
* 3. No standard way to evaluate model assumptions
  + standard assumptions apply (normality, homoscedasticity, linearity, independence) (Kline 2016) when using conventional ML estimation, but you can use asymptotically distribution free (ADF) estimators (Huang & Bentler 2015)
  + when these assumptions are not met, models can be deceiving
    - “…if you assess hypotheses without examining your data, you risk publishing nonsense” (Wilkinson & Task Force on Statistical Inference, 1999, p. 597)
  + while regression-based approaches have visualization tools to evaluate these, the tools are sparse in LVMs
  + problem: estimation is often performed on summary statistics
    - evaluating assumptions requires raw data
    - These summary statistics may have already lost the information required to detect assumption violations
      * *Perhaps come up with an Ambscombs quartet in an SEM setting?*
    - FIML does work on raw data, but assumption evaluation is not a part of the standard SEM toolkit
    - If we had visuals we could highlight problems with model assumptions
* 4. LVM’s are considered “causal models”
  + many statistical models have less lofty goals than LVMs
    - prediction
      * doesn’t matter if the model is “true,” just want to know whether the model can predict future behavior (Reiss 2015; Shmueli 2010)
    - theory-development in regression
      * compare two models of reality
      * goal is only to see which model is better supported (Maxwell & Delaney 2004) – maybe note that this is often done in SEM as well (Gonzalez & Griffin 2001)
  + LVMs: entire ensemble of assumptions “evaluated” simultaneously (Kline 2016)
    - fit is taken as evidence the entirety of the model’s assumptions are true
    - fit is taken as support for a theory
    - problems
      * equivalent/near-equivalent models (Tomarken & Waller 2005)
      * local misfit masked by global fit (Kline 2016)
  + visuals can show which parts of the model are supported by the data and which are not
    - and thus better guide theory
* 5. There is a strong divide between best-practices and actual practices (Goodboy & Kline 2017)
  + people still rely on antiquated rules of thumb poorly supported by empirical evidence
    - as we show in this paper, visuals can highlight problems even when indices suggest the model fits
  + other “rules” people violate? (Tomarken & Waller 2003; 2005)
  + researchers treat “fit” a binary characteristic
    - visuals discourage dichotomization because visuals themselves are continuous
    - visuals are more easily understood than endless lists of rules of thumb
* All these suggest a strong need for intuitive visuals
* Our approach will accomplish several goals
  + highlight global and local misfit
  + provide guidance for correcting misfit
  + assess viability of model assumptions
  + improve encoding through intuitive graphics
  + help applied users better understand what SEM models are doing?
  + all within an easy-to-use R package, flexplavaan, that pairs with lavaan

**Previous Approaches to Visualizing LVMs**

* Hallgren et al. (2019) suggested that researchers move “beyond path diagrams” and begin plotting the statistical models estimated from LVMs.
  + we echo that plea
  + various approaches have been suggested, each with strengths and limitations
* 1. Factor score approaches
  + these approaches create visuals of the latent variables themselves based on the deterministic part of the latent variable model
    - i.e., each dot in a scatterplot is an estimated score
  + Hallgren et al. (2019) utilized scatterplots of latent variables
  + Problems
    - these approaches assume the model actually fits
      * factor scores are only meaningful if the model fits
    - observed variables are not plotted
    - factors are indeterminate
    - not integrated into existing software packages
      * creating visuals is a multistep process (fit the model, extract the factor scores, plot the factor scores)
      * less likely people will utilize these
* 2. Observed variable approaches
  + scatterplot matrices of observed variables with loess lines (Bauer, 2005)
    - designed to identify nonlinearity
    - paired with residual plots of regression of observed nonlinear on observed linear
  + Observed against residuals of observed
    - residuals obtained by regressing observed against observed (Asparouhov & Muthén, 2017; Bauer, 2005)
  + problem:
    - latent variable is not modeled, yet that is often a critical theoretical component of LVMs
    - also not part of any standard software packages
  + Observed against latent (Asparouhov & Muthén, 2017)plotting observed against residuals, implied by LVM (Asparouhov & Muthén, 2017)
    - these seem to detect some misfit but has only sparsely evaluated empirically
      * can it detect misfit besides nonlinearity?
        + missing paths?
        + interaction terms?
        + missing latent variable?
    - integrated in software, but only available in mplus
* In summary: these approaches assume the model fits, fail to model the latent variables, or haven’t been empirically validated for common types of misfit (e.g., missing paths, interaction terms)
  + No common software implementation in the R framework
  + Yet R is commonly used to model LVMs (e.g., through lavaan or OpenMX)

**Our Approach (Linear LVMs)**

**Diagnostic Plots: Trail Plots and Disturbance-Dependence Plots**

**Trail Plots**

* we will utilize standard visualization approaches
* in a typical regression:
  + dots (e.g., in a scatterplot) represents *observed* variables
  + fitted line (e.g., regression) or another symbol (e.g., a large dot to represent the mean in a standard error plot) represent the fit of the model
  + put in different words:
    - regression line is the fit implied by the regression model
* LVM visuals should follow the same convention
  + dots = observed variables
    - e.g., a scatterplot matrix of observed variables as in Bauer (2005)
  + fitted line = fit implied by the model
    - line represents the “trail” left by the unobserved latent variable
    - we call these “trail” plots
* Estimating the fit of the line
  + model-implied correlation \* ratio of standard deviations = slope
  + this line is the fit between the two observed variables, implied by the model
  + if the fit closely resembles the regression line, the model “fits”
    - i.e., it captures the relationship between the two observed variables
* These can diagnose misfit in
  + the measurement model
    - by showing misfit between indicators of the same latent variable
  + the structural model
    - by showing misfit between indicators of different latent variables

**Disturbance-Dependence Plots**

* In regression, a residual dependence plot
  + shows what signal remains after subtracting the fit of the model
  + plot residuals against fitted values
  + after fitting the model there *should* be no signal remaining
  + signal may remain if there’s nonlinear effects (such as polynomial or interaction term)
* Similarly we can create an analogue for LVMs
  + residuals in LVMs refer to the difference between the observed and implied variance/covariance matrix
  + the error in predicting the observed from the model is typically called the “disturbance”
  + Plots of these should thus be called “Disturbance Dependence Plots” (DDPs)
  + with the fit of the model, we can now subtract that from the values of theobserved variables
  + if the model fits well, there should be a flat line

**Measurement Plots**

* Diagnostics tells us whether the estimated factor scores can be believed
* If they are to be believed, we can then develop visuals for the measurement model
* Approach
  + Estimated factor scores on X axis
    - symbol plotted must reflect uncertainty because these are not observed scores
    - we use a line that reflects the upper/lower limits of a 95% confidence or credible interval
  + observed scores on Y axis
    - these have no height (because there’s no uncertainty about their observed scores)
  + to condense visuals, we can convert from wide to long format
    - put observed scores on the same scale
    - panel by the observed variable indicator
* These improve encoding
  + make it apparent which observed variables have highest reliability

**Structural Plots**

* Often in LVMs, the visuals of interest are not the observed, but the latent variables
* Measurement model is ancillary
* As before, we need to reflect uncertainty in the estimate of the latent scores
  + now we show our uncertainty via ellipses
  + uncertainty is estimated using prediction intervals (for frequentist LVMs) or using information from the posterior distribution (for Bayesian LVMs).
* Type of plot will depend on the structural relationships hypothesized
  + e.g., scatterplot for two latent continuous variables, coplots for three latent continuous variables, beeswarm plots for a categorical versus continuous latent variable.
  + for a review, see (Fife, 2019)

**Nonlinear LVMs**

* diagnostic plots will work to *diagnose* nonlinearities
  + will not remedy these
* to remedy nonlinearity, a different approach is needed

**Bayesian LVMs**

* linear LVMs utilize covariance algebra to model fit
* Bayesian LVMs, on the other hand, use raw data
* relationships are specified explicitly
  + as well as residual distributions
* Bayesian LVMs offer several advantages
  + can estimate models impossible to estimate in linear models
  + more flexible (e.g., non-Gaussian latent factors specified in straightforward way)
  + allows users to augment analysis with priors
  + latent scores are estimated as part of the modeling procedure (not a secondary step)
  + missing data handled intuitively
  + after computing factor scores and uncertainty intervals, can compare individual cases more readily than in frequentist
    - e.g., is France (specific case) higher on liberal democracy (latent factor) than Poland?
* blavaan provides an easy-to-use interface for Bayesian LVMs
  + utilizes same syntax as lavaan
  + fitted objects are lavaan objects, so same functions can be utilized for both models
* blavaan generates JAGs or STAN syntax to make it easier to modify MCMC

**Model-Implied Fit For Nonlinear Bayesian LVMs**

* earlier derivation utilized model-implied correlations
  + these assume the fit is actually linear
* if nonlinear, the model-implied fit must be modified
* our approach relies on the estimated factor scores to derive the trail plots
* basic approach
  + model the relationship between the latent variable (x axis) and observed variable (y axis)
    - utilize a smoothed cubic spline function to allow nonlinear patterns
    - store those predictions (call these )
    - do the same for the other variable and store those predictions (call these )
  + serve as the basis of the coordinates of X/Y
    - but are the fit of the model when reliability has been removed
    - it will tend to overestimate the fit of the X/Y relationship
    - must be attenuated proportional to reliability
  + Estimate reliability
    - serves as our best estimate of the true scores of *X*
    - Reliability can be estimated as
  + Attenuate based on reliability
    - Recall that reliability provides upper bound to validity
    - With reliability of zero, the prediction of Y for a given X is the mean of Y
    - With perfect reliability, the relationship between X and Y will be equal to the relationship between
    - To weaken the line, we can adjust the predicted scores to be closer to the mean of *Y*
  + Once we have a model-implied relationship between *X/Y*, we can develop trace plots/DDPs as before

**Examples**

**Well-Fitting Model**

* show a model where rmsea is poor, but the model fits quite well

Omitted Cross-Loading

* show a model (Jedi) where the rmsea is good, but the model fits poorly based on visuals

Nonlinear model

* again, rmsea is good, but the model misses

1. Granted, factor score *estimates* can be observed, but these scores are only estimates, and are only reflective of the latent variables insofar as the model is correct. Additionally, these scores are indeterminate (Rigdon 2019, Steiger 1996). [↑](#footnote-ref-1)