flexplavaan_draft

Contents

Introduction

It is currently an unprecedented time in the social sciences; multiple scientific disciplines are reeling from a "replication crisis" (Camerer et al. 2018; Ioannidis 2005; Pashler and Wagenmakers 2012), new norms for credibility are becoming more prevalent (Nelson, Simmons, & Simonsohn, 2018; Nosek, Ebersole, DeHaven, & Mellor, 2018), and the push for open science is accelerating at a rapid pace (Nosek et al. 2018). Amidst this push for open science practices, some have called for greater use of visualization techniques (Fife and Rodgers 2019; Fife 2020b; Tay et al. 2016). As noted by Tay, et al. (2016), "[visualizations]…can strengthen the quality of research by further increasing the transparency of data…" (p. 694). In other words, one of the best, and most efficient ways of making data analysis open and transparent is to display each and every data point through visualization techniques. This is particularly important in research applications where participant-level data cannot be shared.

Not only do visualizations adhere to the principles of openness and transparency, but they offer several additional advantages; they vastly improve encoding of information (Correll 2015), they highlight model misfit (Healy and Moody 2014), and they are an essential component in evaluating model assumptions (Levine 2018; Tay et al. 2016). As such, we (as well as others, e.g., Fife 2019, 2020b; Wilkinson and Task Force on Statistical Inference 1999) recommend every statistical model ought to be accompanied by a graphic.

Unfortunately, this suggestion is easier said than done. While visualizing some statistical models is trivial (e.g., regressions, *t*-tests, ANOVAs, multiple regression), visualizing others is not. One particularly troublesome class of models to visualize is latent variable models (LVMs). While researcher routinely visualize conceptual models (e.g., via path diagrams), visualizing the statistical models is not so easy. The former visualizations are common, while the latter are not (Hallgren et al. 2019). The reason statistical visualizations of LVM are not intuitive is because they rely on unobserved variables (Bollen 1989). If the variables of interest are unobserved, how can we possibly visualize them?

Though it is not, at first glance, easy to visualize unobserved variables, that does not mean visualizing them is any less important. On the contrary, visualizing latent variables is more important because their presence is unobserved. In the following section, we elaborate on why visualizations are particularly crucial for LVMs. We then review previous approaches others have used for visualizing LVMs, and note their strengths and weaknesses. We then introduce our approach and the corresponding R package `flexplavaan`, which allows users to visualize both lavaan and blavaan objects in R. We then conclude with several examples that highlight how visualizations assisted in identifying appropriate statistical models.

Evaluating Model Fit in LVMs

The validity of LVM-based inferences assume structural models closely approximate real-world causal processes (Bollen 2019; L. Hayduk 2014). Unfortunately, evaluating model adequacy in LVMs is rife with obstacles. For one, misspecifications in any one part of the model can lead to biases that spread throughout the full systems of equations (Bollen 2019). To combat this, SEM practitioners generally rely on global fit tests and approximate fit indices to evaluate the model adequacy (Jackson, Gillaspy Jr, and Purc-Stephenson 2009).

Yet global fit indices themselves represent an obstacle to intelligent model evaluation. Models can yield desirable values (e.g., a non-significant χ^2 test), indicating a strong *overall* model-data correspondence, even when specific aspects of the model are misspecified (Goodboy and Kline 2017; L. Hayduk 2014; Tomarken and Waller 2003). In other words, a nonsensical model can still yield estimates that lead one to believe in their own statistical models. Moreover, in our experience, applied users often lack an intuitive understanding of what global fit statistics tell them about their models. Does a TLI of 0.95 mean we have established a strong theoretical foundation for a model? If RMSEA dips below 0.05, should we consider our model statistically significant?

While global fit measures may be useful in identifying problematic model features, users instead rely on conventional cutoffs (e.g, Hu and Bentler 1998) to determine whether a model is "adequate," a practice that has received resounding criticism (Barrett 2007; Chen et al. 2008; L. A. Hayduk 2014; McIntosh 2007). This lack of understanding is evident when applied users express confusion about why global fit indices provide conflicting assessments of model fit (Lai and Green 2016).

A number of scholars have emphasized the importance of supplementing SEM global fit indices with local fit assessment, investigating the tenability of all specific model implications individually (Bollen 2019; Goodboy and Kline 2017; **???**; Thoemmes, Rosseel, and Textor 2018; Tomarken and Waller 2003, 2005). Local fit evaluation procedures (e.g., inspection of residual correlation matrices (Bollen 1989), confirmatory tetrad analysis (Hipp and Bollen 2003), and equation-based overidentification tests (Bollen 2019)) can help identify individual model specifications that are inconsistent with the data and may give clues about effective remedial strategies (Bollen 1989; Goodboy and Kline 2017; Tomarken and Waller 2003, 2005).

While local fit indices will certainly improve model evaluation, they too suffer from a number of problems. First, [steve… care to add some here].

One problem that both global and local fit indices share is that they are a highly *compressed* representations of both the data and the model. Many readers may be familiar with Amscombe's quartet (reproduced in Figure xx, xxxx). There is a many-to-one relationship between data and indices; very different types of data may yield identical fit indices. Some of these data patterns may be very poorly represented by the model. This problem is only exacerbated with LVMs simply because traditional algorithms compress the data at multiple levels (e.g., raw data are compressed into means/covariances, which are then compressed into model parameter estimates and/or indices that evaluate model fit). Put differently (and perhaps quite cynically), LVM is the process of compressing hundreds or thousands of datapoints into a handful (or less) of estimates that may or may not represent the data-generating process. When considered from this perspective, the most common practices of evaluating model fit seem primitive, at best.

The best defense against this compression is to rely on modeling evaluation strategies that evaluate uncompressed data. Perhaps the best (if not only) way to evaluate uncompressed data is through visualizations, particularly visuals that display raw data (Fife 2020b).

Previous Approaches to Visualizing LVMs

Ironically, while visuals of statistical models are both intuitive and effective at highlighting model misfit, there is sparse literature describing visual strategies for evaluating the adequacy of SEMs, at least when compared to the relative to the extensive literature discussing the merits of global fit tests and indices (e.g., Barrett 2007; Chen et al. 2008; L. A. Hayduk 2014; Hu and Bentler 1998; McIntosh 2007; Shi and Maydeu-Olivares 2020; Smith and McMillan 2001; Steiger 2007). While reporting of numeric fit indices is ubiquitous in LVM applications (Jackson, Gillaspy Jr, and Purc-Stephenson 2009), plots of data underlying LVMs are rare (Hallgren et al. 2019). There are, however, some strategies for using visuals to evaluate the tenability of model assumptions, diagnose causal misspecifications, and select the best model from a group of competitors.

A common visual approach for identifying model-data discrepancies is to plot the distribution of the residuals of the covariances/correlation matrix (e.g., using stem-and-leaf plots or histograms) (Bollen 1989). This can aid in identifying specific components of the data that the model struggles to capture (Bollen 1989; Bollen and Arminger 1991), which might not be detected with global fit indices (Goodboy and Kline 2017; Tomarken and Waller 2003, 2005).

Unfortunately, these plots suffer from two major problems. First, the residuals in this case are themselves *compressed* estimates. As such, we might have a model that is poorly represented by linear correlation (e.g., if the data contain nonlinear relationships), but that problem would never be uncovered by studying residual plots.

A second problem with plots of correlation/covariance residuals is they are extremely limited in the amounts of misspecification they might reveal. For example, suppose we model a latent variable with three indicators (x_1, x_2, and x_3), but the data-generating model actually has x_3 associated with x_2/x_1, but not an indicator of the latent variable (see Figure 1). Depending on the other variables in the model, these two models will have the same implied variance/covariance matrix.[1]

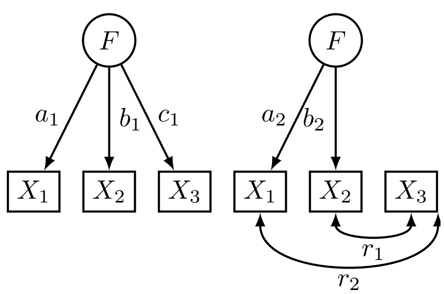


Figure 1: The model on the left is the user-specified model, while the model on the right is the data-generating model. These two models will yield very different estimates of a/b, but have the same implied correlation between x_1 and x_2.

Global fit indices and residual covariances, on account of their focus on aggregate statistics, do not identify aberrant cases that may influence model parameter estimates and fit (Bollen and Arminger 1991). Bollen and Arminger (1991) demonstrated methods of calculating raw and standardized individual case residuals (ICRs) representing differences between observed and model-estimated case values for outcome variables (Bollen and Arminger 1991). They used stem-and-leaf, index, and histogram plots to help locate outlying and influential observations. Pek and MacCallum (2011) demonstrated how case diagnostic procedures commonly used in generalized linear models (e.g., Mahalanobis distance, generalized Cook's D, and DFBETAs) can

[1] In the left model, the standardized relationship between x_1/x_3 is $a_1 \times c_1$, while in the right model it is r_2. Likewise, the x_1/x_2 relationship is $a_1 \times b_1$ in the left model, while it is r_1 in the right model. The LVM machinery will attempt to make $a_1 \times c_1 = r(x_1, x_3)$, which is the same as setting $a_1 \times c_1 = r_2$ (and it will set $a_1 \times b_1$ to r_1).

be applied to SEMs to detect influential cases with index plots clearly identifying the most aberrant cases. Flora, LaBrish, and Chalmers (2012) applied these diagnostic procedures and others specifically to factor analysis models, and Yuan and Hayashi (2010) used visualizations of Mahalanobis distance metrics to identify high-leverage cases and outliers. Visualization procedures showing case influence on model fit (e.g., likelihood differences) and parameter estimations (e.g., generalized Cook's D) have been implemented in open-source R packages, including faoutlier (Chalmers 2017) and influence.SEM (Pastore and Pastore 2018).

Asparouhov and Muthén (2017) used estimated factor scores and ICRs to detect specific structural misspecifications. First, they showed that plots of estimated factor scores for a latent outcome variable against observed predictor variables can be used to detect unspecified nonlinear effects of the predictor on the latent outcome. Second, they used ICR scatterplots to detect violations of local independence in a congeneric latent factor model. Finally, they demonstrated in a latent factor model how plotting predicted values for a reflective indicator against the observed indicator values could aid in uncovering unmodeled heterogeneity that could be better captured using a mixture model.

Raykov and Penev (2014) used visualizations of ICRs to aid in model selection in the context of latent growth curve modeling. When comparing linear and quadratic growth curve models for the same data, for example, they showed that a scatterplot of the ICRs for the quadratic model vs. ICRs for a linear model can help identify which model best minimizes model-data discrepancies. In the context of growth mixture modeling, Wang, Hendricks Brown, and Bandeen-Roche (2005) showed how visualization of empirical Bayes residuals (e.g., Q-Q and trajectory plots) can aid in determining the appropriate number of classes, an adequate shape of within-class growth trajectories, and missing confounders.

A potential limitation of visual model-evaluation procedures based on ICRs is that they rely on factor score estimates. Individual latent factor scores cannot be uniquely determined (Grice 2001; Rigdon, Becker, and Sarstedt 2019; Steiger 1996). In cases where factors are highly indeterminate – e.g., factors with few indicators only weakly predicted by the latent factor – different factor score estimation methods can yield highly discrepant values, potentially even estimates that are negatively correlated (Grice 2001). Therefore, in the presence of high levels of factor indeterminacy, the conclusions drawn from visual diagnostic strategies using individual scores could vary depending on which factor score estimation strategy is employed.

Our Approach (Linear LVMs)

Diagnostic Plots: Trail Plots

To begin how to conceptualize LVMs, let us first consider how typical linear models are visualized. In a standard regression, each dot in a scatterplot represents scores on the observed variables. Often, analysts overlay additional symbols to represent the fit of the model (e.g., a line to represented the fitted regression model, or large dots to represent the mean). Sometimes additional symbols are overlaid to represent uncertainty (e.g., confidence bands for a regression line or standard error bars). See Figure 2 as an example. In either case, the dots represent observed information, while the fitted information is conveyed using other symbols.

Likewise, visualizing LVMs ought to follow similar conventions; the dots should represent the observed information, as in Bauer (2005). In his visuals, pairwise relationships between observed variables are represented in a scatterplot. However, Bauer's approach did not overlay a model-implied fit, as we seek to do. When the line represents the model-implied fit, it denotes the trail left behind by the unobserved latent variable. As such, we call these plots "trail plots." How then does one identify the slope/intercept of the LVM's model-implied fit? It is quite easy to do so when standard linear LVMs are used. Suppose we have a factor (T) with three indicators (e.g., X_1, X_2 and X_3), and we wish to visualize the pairwise trace plot between X_1 and X_2. To do so, we can simply utilize the model-implied correlation matrix:

$$\beta_{x_1|x_2} = \hat{r}_{x_1,x_2} \frac{s_{x_1}}{s_{x_2}}$$

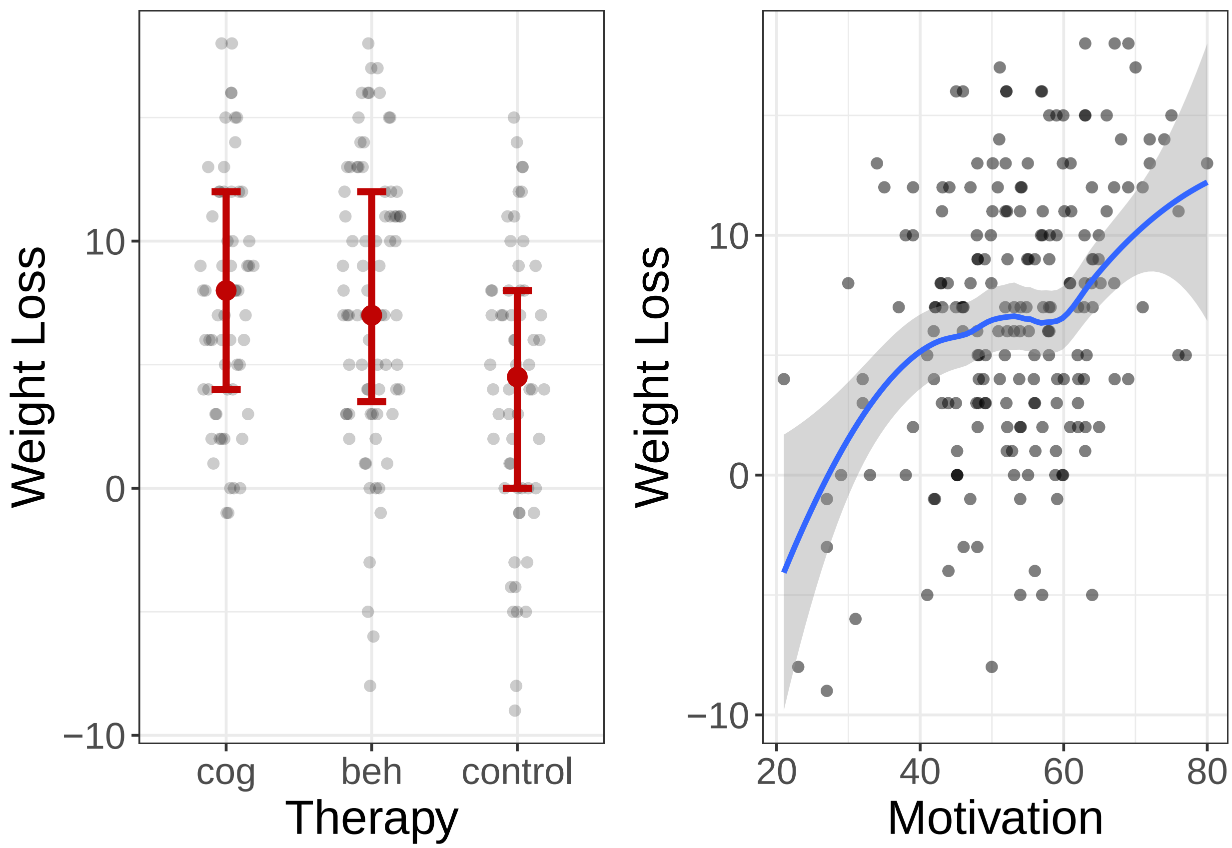


Figure 2: Example figure that shows how standard statistical models are visualized. Dots represent scores on observed variables, while other symbols (e.g., regression line, large dots) represent the fit of the model.

where $\hat{r}_(x_1, x_2)$ is the model-implied correlation between X_1 and X_2, s_{x_1} and s_{x_2} are the standard deviations of the two variables. One can then estimate the intercept using basic algebra:

$$b_0 = \bar{X}_1 - \beta_(x_1|x_2)\bar{X}_2$$

Figure 3 shows the LVM model-implied fit in red with a regression line in blue for simulated data. Because the regression line minimizes the sum of squared errors, we would hope that the LVM fitted line (red) closely approximate the regression line. In this case, the two overlap quite extensively. On the other hand, if the two lines differ, we can be certain the LVM fails to capture the entire relationship between the two observed variables.

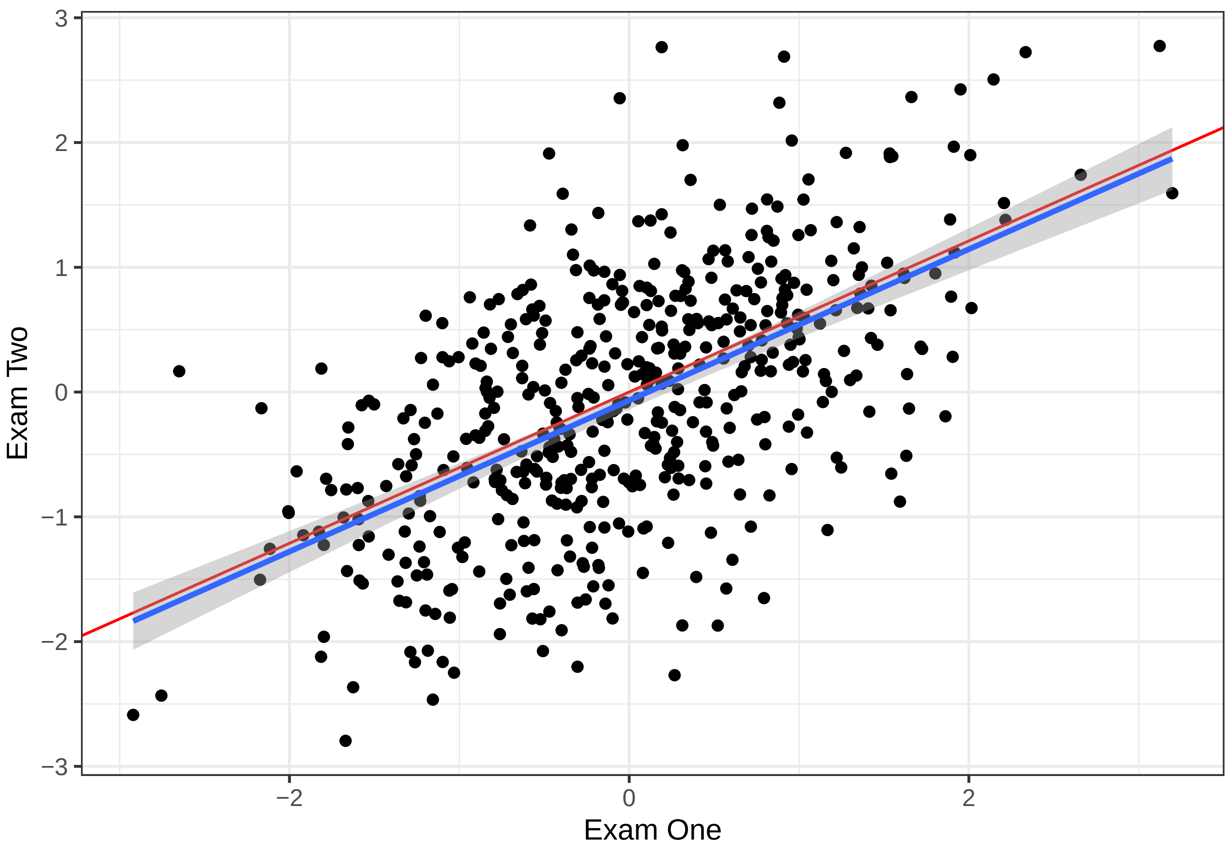


Figure 3: The LVM-implied fit between X1 and X2, shown in red. The blue line represents the regression line between the two variables. The more closely the model-implied fit line resembles the regresson line, the better the fit of the LVM.

Of course, Figure 3 only shows one pairwise relationship between variables. If we wished to visualize all the variables in our model, we would have to utilize a scatterplot matrix, as in Figure 4. Naturally, this becomes quite cumbersome when users have more than seven or eight variables. In this case, it is best to visualize only a subset of variables. We will later discuss strategies for how best to select appropriate subsets.

The primary advantage of trail plots is that they easily show misfit in LVMs. They do so by showing differences in fit between indicators of different variables. For example, 5 shows a model where $x1 - x2$ load onto one latent variable, x4-x5 load onto another, and x3 loads onto both; however, the specified model assumes x3 only loads onto the first model.

Another advantage of trail plots is they visually (and often times strikingly) show how little information a model might capture. For example, Figure 6 shows a model that, by the fit indices, has remarkably good fit:

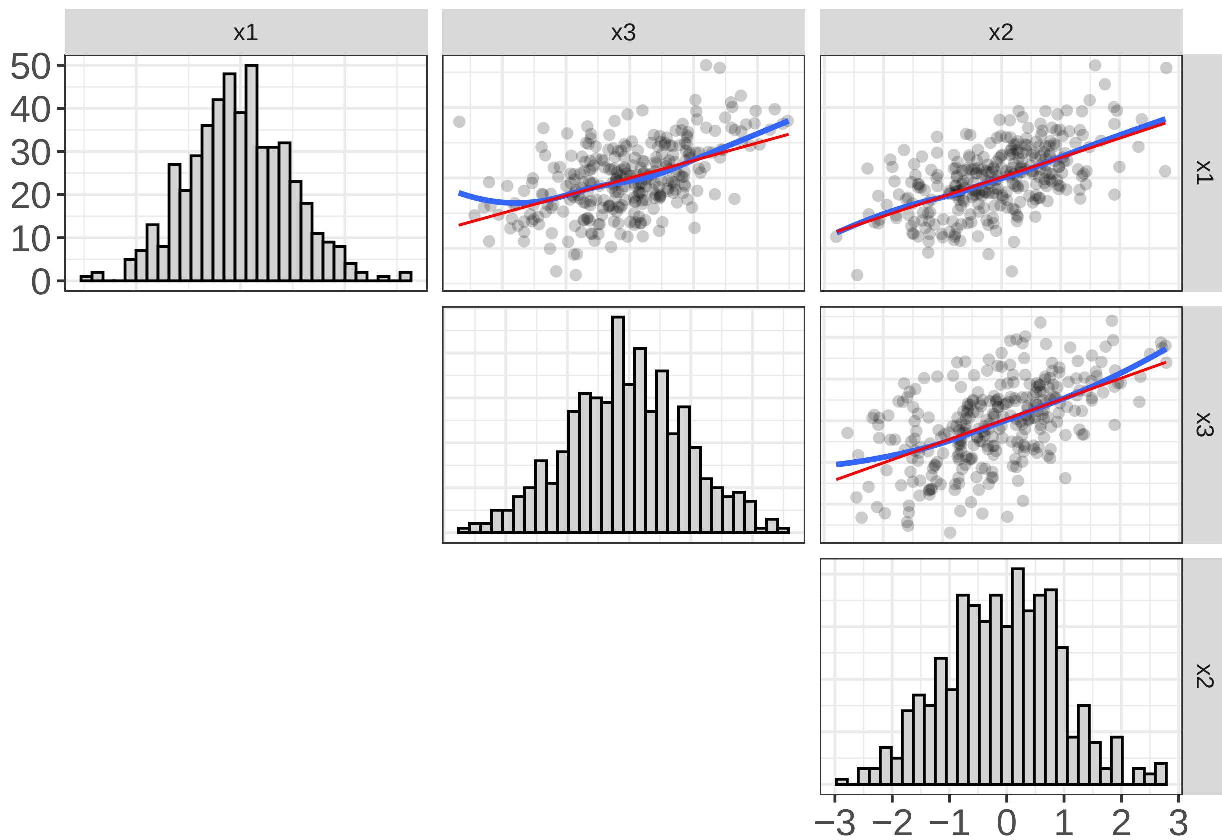


Figure 4: Scatterplot matrix showing the model-implied fit (red) and regression-implied fit (blue) between three simulated indicator variables. The diagonals show the histograms.

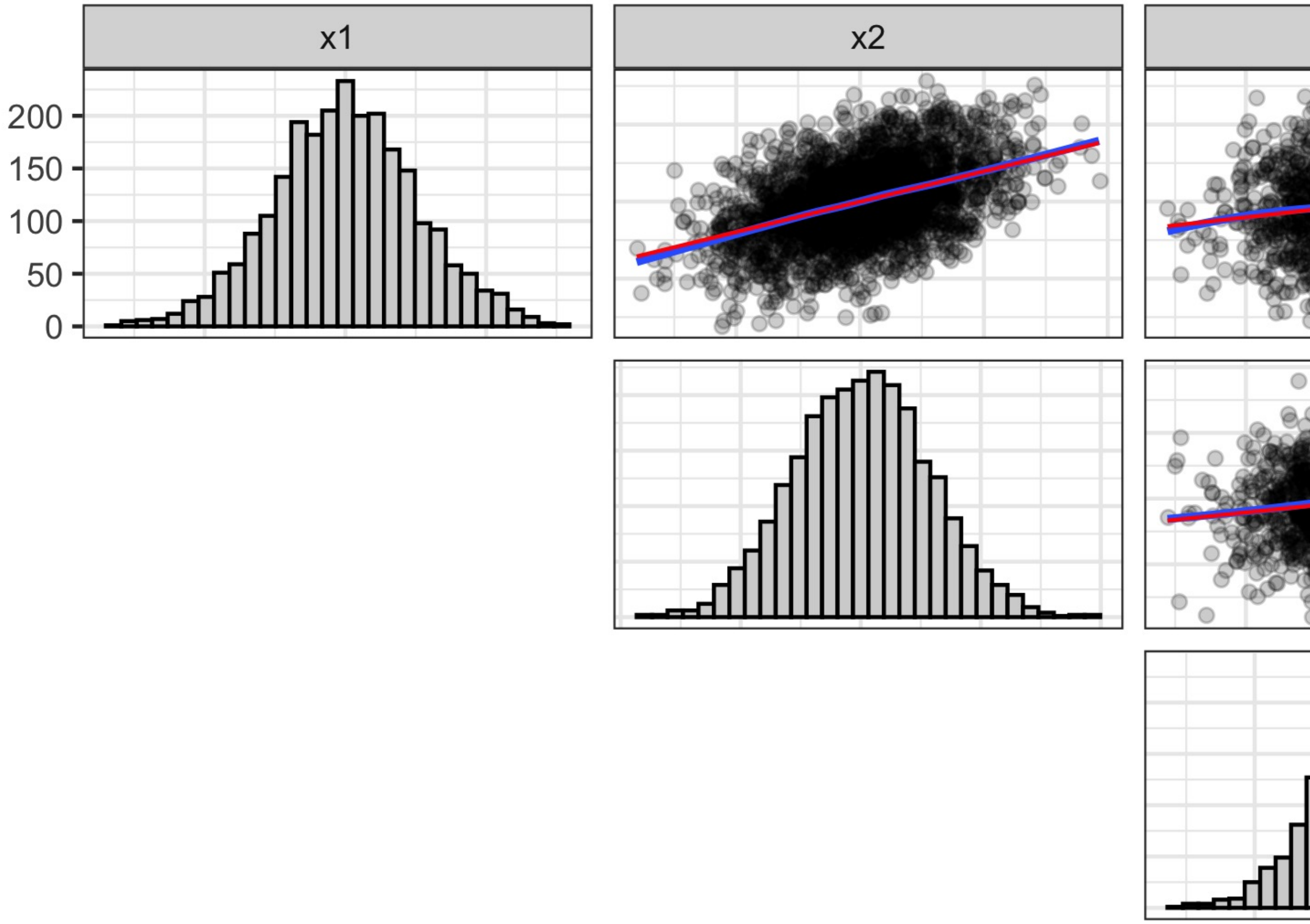


Figure 5: This plot shows structural misfit; x3 loads onto two factors, but only one is modeled.

RMSEA is <0.001, TLI/CFI/NNFI are all approximately 1, and RMR is approximately zero. However, the visual shows the model actually captures very little information; while the two lines are quite similar, the majority of slopes are very near flat.

Disturbance-Dependence Plots

One common technique for visualizing the adequacy of statistical models in classic regression is residual-dependence plots. With these graphics, one simply plots the residuals of the model (Y axis) against the predicted values (X axis). The rationale behind this is simple: the model should have extracted any association between the prediction and the outcome. The residuals represent the remaining information after extracting the signal from the model. If there is a clear trend remaining in the data (e.g., a nonlinear pattern or a "megaphone" shape in the residuals), this indicates the model failed to capture important information.

Likewise, in LVMs, we can apply this same idea to determine whether the fit implied by the LVM has successfully extracted any association between any pair of predictors. However, in LVMs, residuals refer to the discrepancy between the model-implied and the actual variance/covariance matrix (or correlation matrix). As such, naming these plots "residual-dependence plots" would be a misnomer. Rather, misfit at the raw data level is typically called a disturbance. As such, we call these plots disturbance-dependence plots.

Like trace plots, we visualize disturbance-dependence plots for each pair of observed variables. To do so, `flexplavaan` subtracts the fit implied by the trace plots from the observed scores. For example, a disturbance dependence plot for an X_1/X_2 relationship would subtract the "fit" of X_2 implied by the trace plot from the actual X_2 scores (and vice versa for the X_2/X_1 relationship). If the trace-plot fit actually extracts all association between the pair of observed variables, we would expect to see a scatterplot that shows no remaining association between the two. If there is a pattern in the scatterplot remaining, we know the fit of the model misses important information about that specific relationship. To aid in interpreting these plots, we can overlay the plot with a flat line (with a slope of zero), as well as a regression (or loess) line. The first line indicates what signal should remain after fitting the model, while the second line shows what actually remains. Figure 7 shows an example of trace plots in the upper triangle and disturbance-dependence plots in the lower triangle of a scatterplot matrix. These plots are for the same data shown in the right image of Figure 5. Notice how the plots associated with X_3 all have positive slopes, indicating the model failed to capture important signal remaining in the model.

Together, both of these plots (trace plots and diagnostic-dependence plots) serve as a critical diagnostic check. Both these plots will signal misfit both in the measurement and structural components of the model. Conversely, these models will also help users determine whether the model is to be believed. If they show the model adequately fits the data, the user can then proceed to plot two different types of plots: measurement plots and structural plots.

Measurement Plots

One of the primary purposes of the diagnostics is to determine whether one's conceptualization of the latent variables is to be believed. If the trace plots and disturbance dependence plots indicate the LVM is a good representation of the data, one can be more confident the latent variables are properly estimated. If this is the case, we can now make a step toward visualizing the latent variables themselves.

The approach we suggest is to plot the factor scores on the Y axis and the observed scores on the X axis. Naturally, this means one could really only visualize one indicator at a time, which is a serious limitation for most (if not all) latent variable models. To overcome this problem, we recommend paneling each indicator, as in Figure 8. To do so, this requires converting our data from "wide" to "long" format (see Table 2), meaning that variables that once occupied separate columns (e.g., $x1, x2,$ and $x3$; see Table 1) are now collapsed into two columns, one of which contains the values of $x1 - x3$, the other of which indicates which variable the

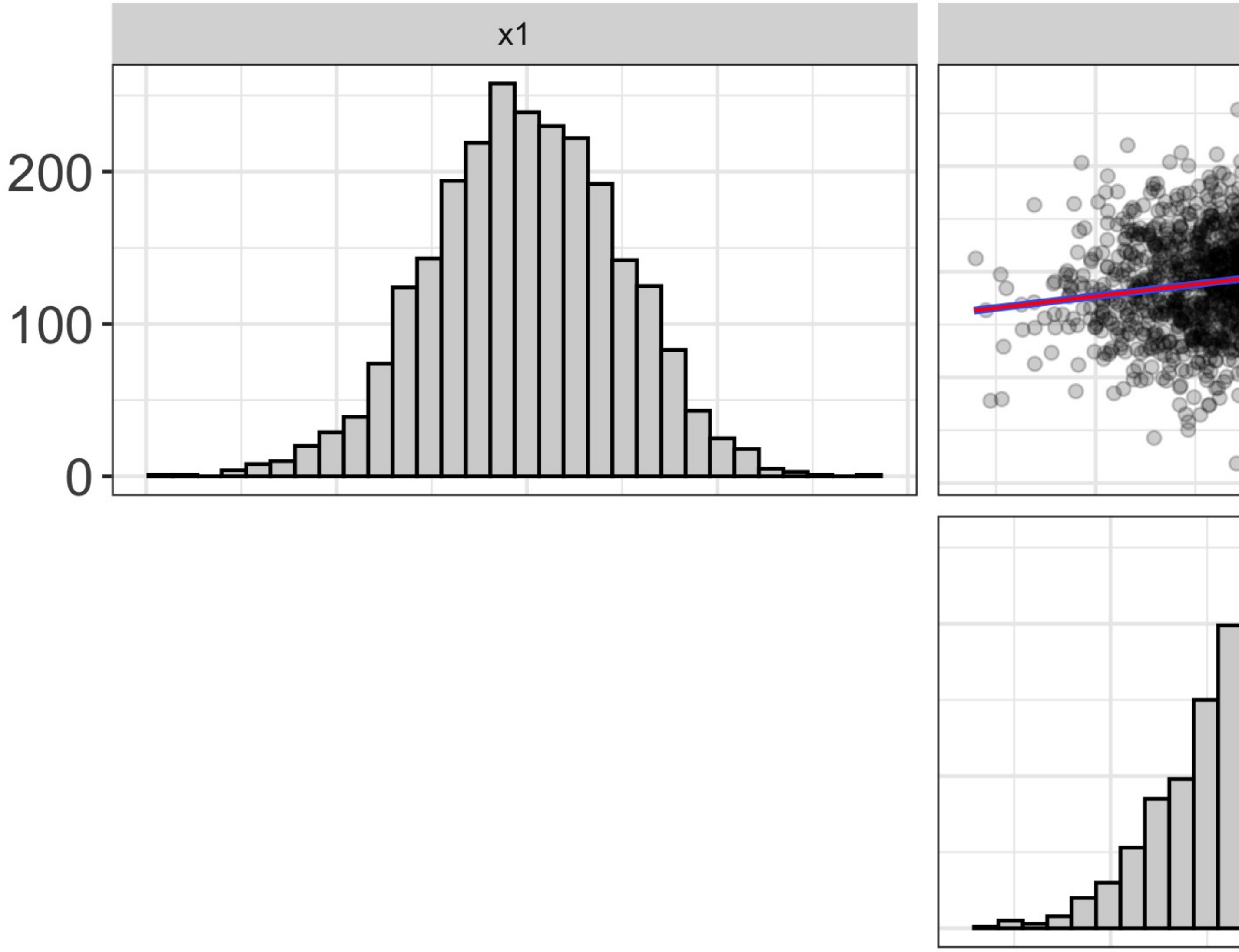


Figure 6: This plot shows a model that, by standard fit indices, fits quite well. The visuals, on the other hand, illustrate that the model isn't capturing

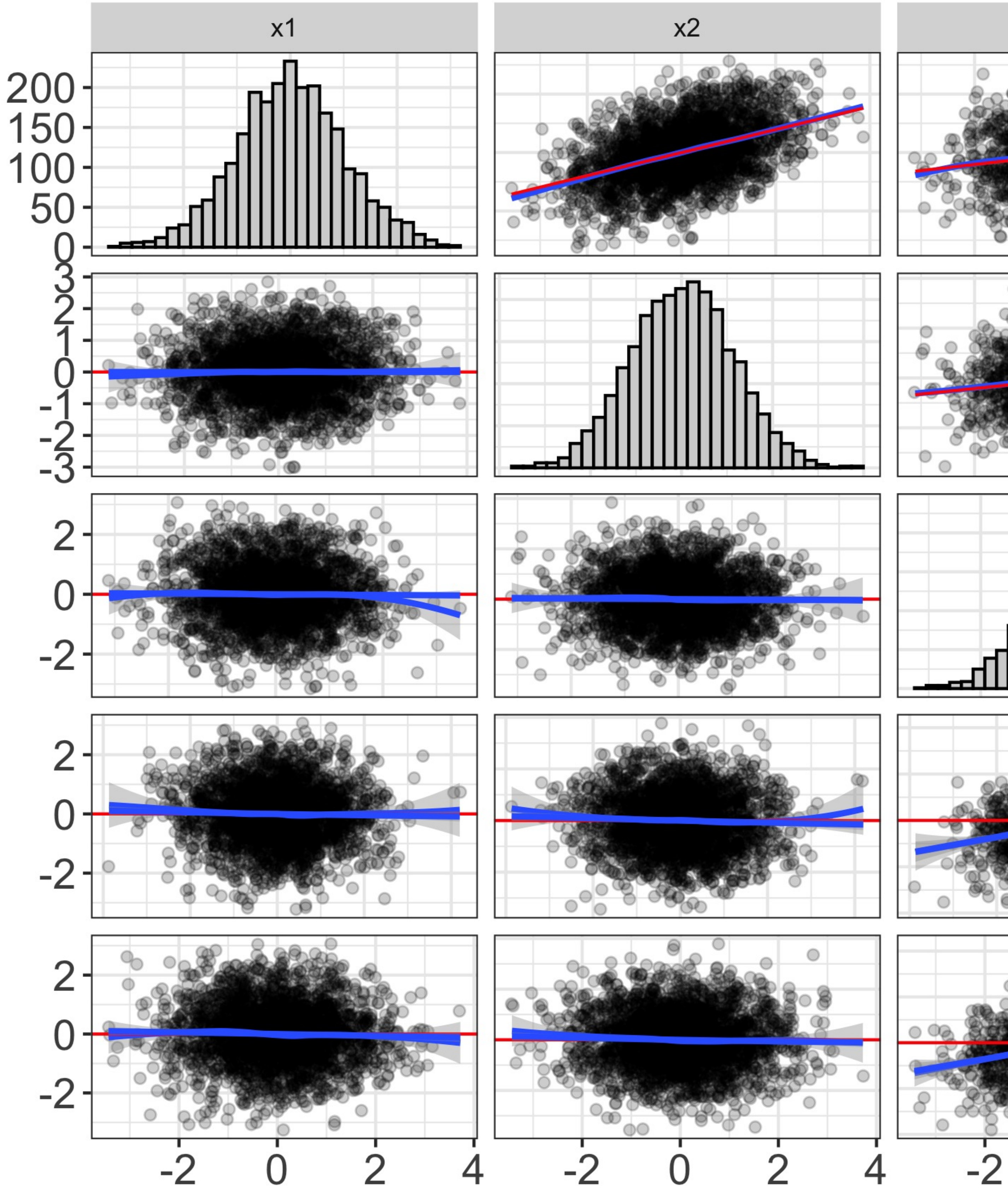


Figure 7: This plot shows a disturbance-dependence plot for the same data visualized in Figure 5 in the lower triangle.

Table 1: A Table of Simulated Data Showing the Data in Wide Format

x1	x2	x3
8	9	9
8	9	9
9	11	9

Table 2: This Table Contains the Same Data in Table 1, but standardized and in “Long” Rather Than “Wide” Format.

Measure	Observed
x1	-0.5773503
x1	-0.5773503
x1	1.1547005
x2	-0.5773503
x2	-0.5773503
x2	1.1547005
x3	NaN
x3	NaN
x3	NaN

measurement belongs to (see Table 2). We should also convert each variable to z-scores to make it easier to compare the observed/latent relationship across indicators, as in Table 2.

Another alteration from a standard scatterplot is the use of vertical bars to indicate uncertainty in estimating the latent variables. In Figure **??**, the dots represent the estimated factor scores and the lines indicate $\pm 1SD$. The red line is a “ghost line” (Fife 2020a), which simply repeats the line from the second panel ($x2$) to the other panels. This makes it easier to identify which indicators best load onto the latent variable (i.e., which indicator is most reliable). In Figure 8, $x2$ has the largest factor score.

Naturally, there is a great deal of overlapping points/lines in Figure 8. To minimize overlap, we can instead sample datapoints, as in Figure **??**.

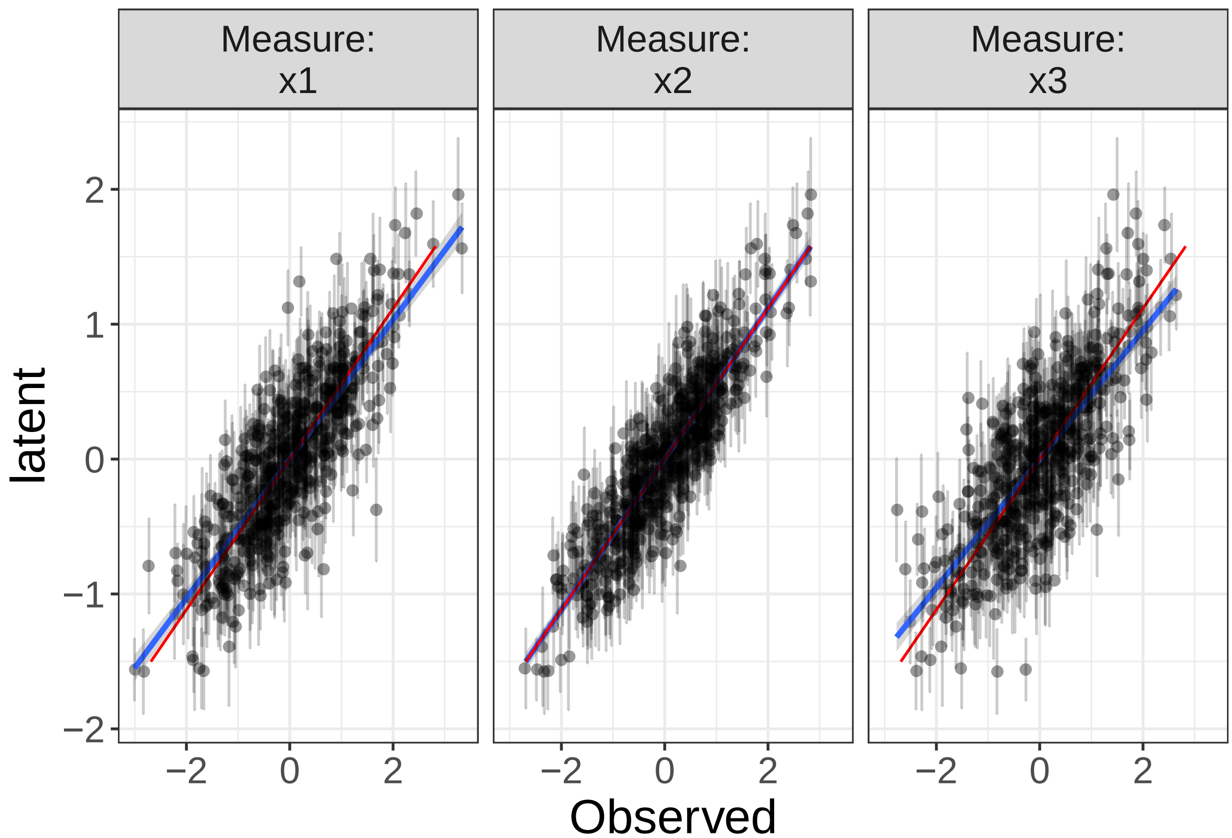
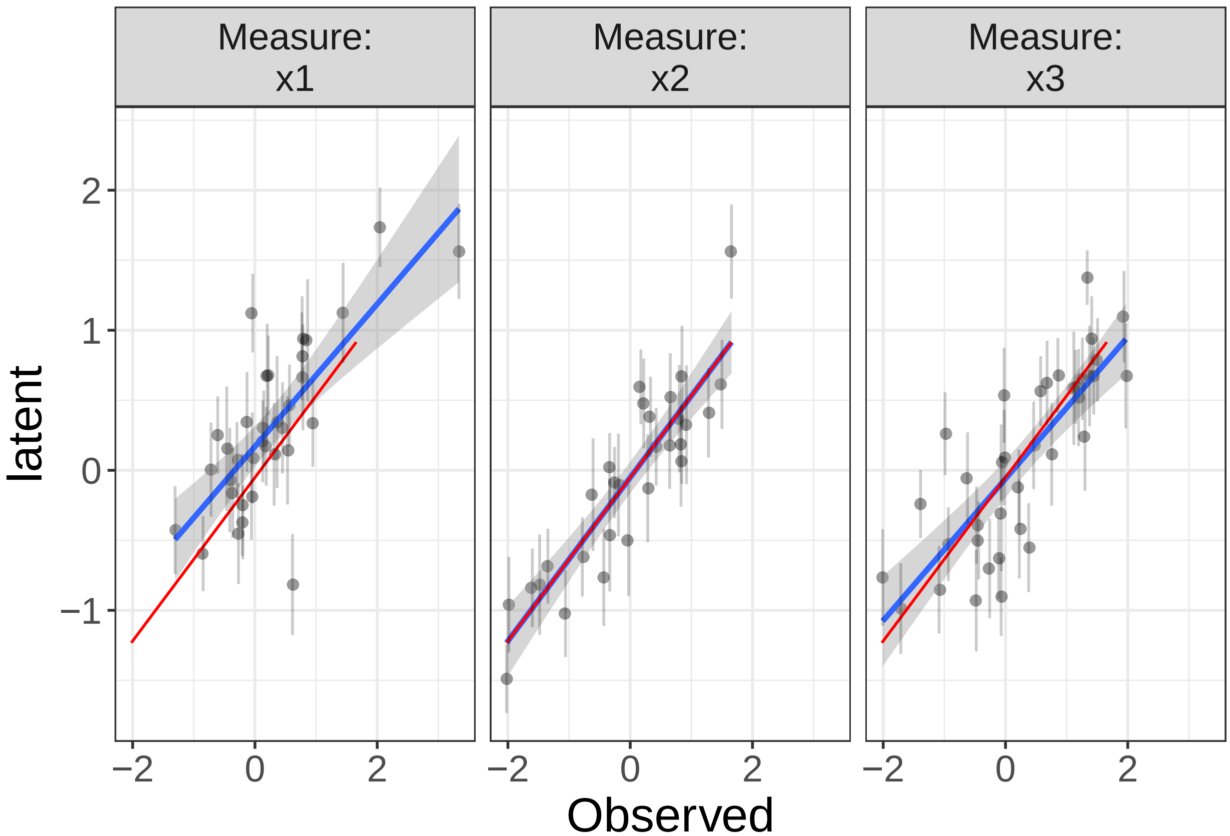


Figure 8: This image, called a measurement plot, shows the relationship between the latent variable (Y axis) and each of the standarized indicators ($x1, x2$, and $x3$).



Structural (Beech Ball) Plots

When modeling latent variables, often the visuals of interest are not the observed variables, but the latent variables. In other words, the measurement model is ancillary to the substantive model. Naturally, we might wish to visualize the relationship between the latent variables.

However, as before, we need to have visuals that reflect uncertainty in our estimates of the latent scores. With measurement plots, we needed only to reflect that uncertainty on the Y axis (because the Y axis displayed the latent variable while the X axes dispalyed the observed variables). When visualizing relationships between latent variables, on the other hand, both axes should reflect uncertainty. In `flexplavaan`, this uncertainty is represented as ellipses. The diameter of the ellipses (for both the X axis and the Y axis) are obtained from prediction intervals for `lavaan` objects or from posterior distributions for `blavaan` objects. Figure 9 shows these plots, which we call "Structural Plots," or "Beech Ball Plots" (because the ellipses look like beech balls in various stages of compression).

Aside from the beech balls, there is a great deal of flexibility in how one visualizes the structural model. In our example, we only had two variables to visualize, so a simple bivariate plot was most natural. When more variables are included, we might utilize paneling, added variable plots, beeswarm plots, etc. For a review of the types of plots possible, see Fife (2020a).

References

Asparouhov, Tihomir, and Bengt Muthén. 2017. "Using Mplus individual residual plots for diagnostics and model evaluation in SEM."

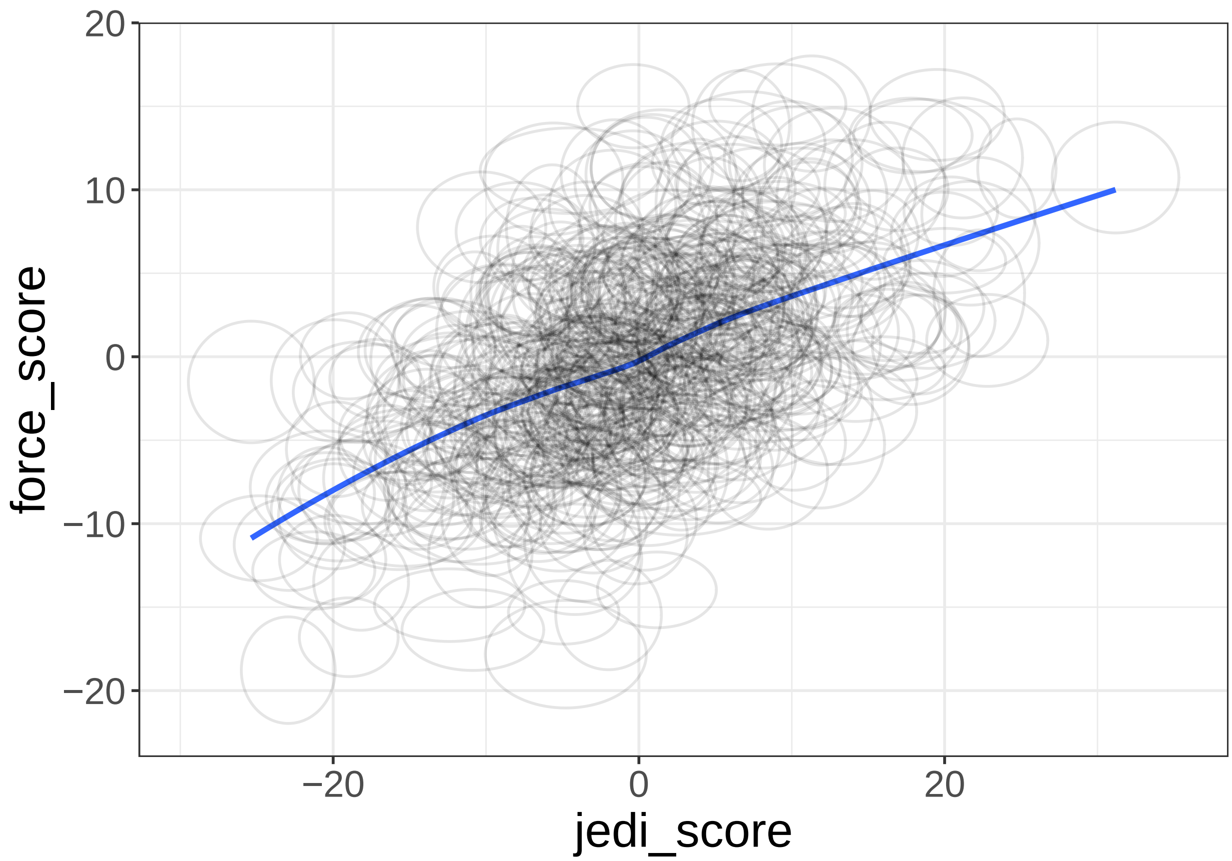


Figure 9: Structural or “Beech Ball” plot of the relationship between the latent variables Force and Jedi. The ellipses represent the prediction intervals for the factor scores of the latent variables.

Barrett, Paul. 2007. "Structural Equation Modelling: Adjudging Model Fit." *Personality and Individual Differences* 42 (5): 815–24.

Bauer, Daniel J. 2005. "The role of nonlinear factor-to-indicator relationships in tests of measurement equivalence." *Psychological Methods* 10 (3): 305–16. https://doi.org/10.1037/1082-989X.10.3.305.

Bollen, Kenneth A. 1989. *Structural equations with latent variables.* Wiley Series in Probability and Mathematical Statistics. {Applied} Probability and Statistics Section; 0271-6356. John Wiley & Sons.

———. 2019. "Model Implied Instrumental Variables (Miivs): An Alternative Orientation to Structural Equation Modeling." *Multivariate Behavioral Research* 54 (1): 31–46.

Bollen, Kenneth A, and Gerhard Arminger. 1991. "Observational Residuals in Factor Analysis and Structural Equation Models." *Sociological Methodology*, 235–62.

Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2018. "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015." https://doi.org/10.1038/s41562-018-0399-z.

Chalmers, Phil. 2017. "Package 'Faoutlier'."

Chen, Feinian, Patrick J. Curran, Kenneth A. Bollen, James Kirby, and Pamela Paxton. 2008. "An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models." *Sociological Methods and Research* 36 (4): 462–94. https://doi.org/10.1177/0049124108314720.

Correll, Michael A. 2015. "Visual Statistics." Doctoral Dissertation, University of Wisconsin-Madison.

Fife, Dustin A. 2019. "A Graphic is Worth a Thousand Test Statistics: Mapping Visuals onto Common Analyses." http://rpubs.com/dustinfife/528244.

———. 2020a. "Flexplot: Graphical-Based Data Analysis." *PsyArxiv.* https://doi.org/10.31234/osf.io/kh9c3.

———. 2020b. "The Eight Steps of Data Analysis: A Graphical Framework to Promote Sound Statistical Analysis." *Perspectives on Psychological Science* 15 (4): 1054–75. https://doi.org/10.1177/1745691620917333.

Fife, Dustin A., and Joseph Lee Rodgers. 2019. "Exonerating EDA: Addressing the Replication Crisis By Expanding the EDA/CDA Continuum." *Unpublished Manuscript.* http://quantpsych.net/fife-exonerating-eda-draft-oct2019-df-edits/.

Flora, David B, Cathy LaBrish, and R Philip Chalmers. 2012. "Old and New Ideas for Data Screening and Assumption Testing for Exploratory and Confirmatory Factor Analysis." *Frontiers in Psychology* 3: 55.

Goodboy, Alan K., and Rex B. Kline. 2017. "Statistical and Practical Concerns With Published Communication Research Featuring Structural Equation Modeling." *Communication Research Reports* 34 (1): 68–77. https://doi.org/10.1080/08824096.2016.1214121.

Grice, James W. 2001. "Computing and Evaluating Factor Scores." *Psychological Methods* 6 (4): 430.

Hallgren, Kevin A., Connor J. McCabe, Kevin M. King, and David C. Atkins. 2019. "Beyond path diagrams: Enhancing applied structural equation modeling research through data visualization." *Addictive Behaviors* 94 (March 2018): 74–82. https://doi.org/10.1016/j.addbeh.2018.08.030.

Hayduk, Leslie. 2014. "Seeing Perfectly Fitting Factor Models That Are Causally Misspecified: Understanding That Close-Fitting Models Can Be Worse." *Educational and Psychological Measurement* 74 (6): 905–26. https://doi.org/10.1177/0013164414527449.

Hayduk, Leslie A. 2014. "Shame for Disrespecting Evidence: The Personal Consequences of Insufficient Respect for Structural Equation Model Testing." *BMC Medical Research Methodology* 14 (1): 124.

Healy, Kieran, and James Moody. 2014. "Data Visualization in Sociology." *Annual Review of Sociology* 40 (1): 105–28. https://doi.org/10.1146/ANNUREV-SOC-071312-145551.

Hipp, John R, and Kenneth A Bollen. 2003. "Model Fit in Structural Equation Models with Censored, Ordinal, and Dichotomous Variables: Testing Vanishing Tetrads." *Sociological Methodology* 33 (1): 267–305.

Hu, Li-tze, and Peter M Bentler. 1998. "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification." *Psychological Methods* 3 (4): 424.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

Jackson, Dennis L, J Arthur Gillaspy Jr, and Rebecca Purc-Stephenson. 2009. "Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations." *Psychological Methods* 14 (1): 6.

Lai, Keke, and Samuel B Green. 2016. "The Problem with Having Two Watches: Assessment of Fit When Rmsea and Cfi Disagree." *Multivariate Behavioral Research* 51 (2-3): 220–39.

Levine, Sheen S. 2018. "Show us your data: Connect the dots, improve science." *Management and Organization Review* 14 (2): 433–37. https://doi.org/10.1017/mor.2018.19.

McIntosh, Cameron N. 2007. "Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007)." *Personality and Individual Differences* 42 (5): 859–67. https://doi.org/10.1016/j.paid.2006.09.020.

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. "The preregistration revolution." *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1708274114.

Pashler, Harold, and Eric-Jan Wagenmakers. 2012. "Editors' Introduction to the Special Section on Replicability in Psychological Science." *Perspectives on Psychological Science* 7 (6): 528–30. https://doi.org/10.1177/1745691612465253.

Pastore, Massimiliano, and Maintainer Massimiliano Pastore. 2018. "Package 'Influence. SEM'."

Pek, Jolynn, and Robert C MacCallum. 2011. "Sensitivity Analysis in Structural Equation Models: Cases and Their Influence." *Multivariate Behavioral Research* 46 (2): 202–28.

Raykov, Tenko, and Spiridon Penev. 2014. "Latent Growth Curve Model Selection: The Potential of Individual Case Residuals." *Structural Equation Modeling: A Multidisciplinary Journal* 21 (1): 20–30.

Rigdon, Edward E, Jan-Michael Becker, and Marko Sarstedt. 2019. "Factor Indeterminacy as Metrological Uncertainty: Implications for Advancing Psychological Measurement." *Multivariate Behavioral Research* 54 (3): 429–43.

Shi, Dexin, and Alberto Maydeu-Olivares. 2020. "The Effect of Estimation Methods on Sem Fit Indices." *Educational and Psychological Measurement* 80 (3): 421–45.

Smith, Tracy D, and Bradley F McMillan. 2001. "A Primer of Model Fit Indices in Structural Equation Modeling."

Steiger, James H. 2007. "Understanding the limitations of global fit assessment in structural equation modeling." *Personality and Individual Differences* 42 (5): 893–98. https://doi.org/10.1016/j.paid.2006.09.017.

Steiger, James H. 1996. "Dispelling Some Myths About Factor Indeterminancy." *Multivariate Behavioral Research* 31 (4): 539–50.

Tay, Louis, Scott Parrigon, Qiming Huang, and James M LeBreton. 2016. "Graphical Descriptives: A Way to Improve Data Transparency and Methodological Rigor in Psychology." *Perspectives on Psychological Science* 11 (5): 692–701. https://doi.org/10.1177/1745691616663875.

Thoemmes, Felix, Yves Rosseel, and Johannes Textor. 2018. "Local fit evaluation of structural equation models using graphical criteria." *Psychological Methods* 23 (1): 27–41. https://doi.org/10.1037/met0000147.

Tomarken, Andrew J, and Niels G Waller. 2003. "Potential Problems with" Well Fitting" Models." *Journal of Abnormal Psychology* 112 (4): 578.

Tomarken, Andrew J., and Niels G. Waller. 2005. "Structural Equation Modeling: Strengths, Limitations, and Misconceptions." *Annual Review of Clinical Psychology* 1 (1): 31–65. https://doi.org/10.1146/annurev.clinpsy.1.102803.144239.

Wang, Chen-Pin, C Hendricks Brown, and Karen Bandeen-Roche. 2005. "Residual Diagnostics for Growth Mixture Models: Examining the Impact of a Preventive Intervention on Multiple Trajectories of Aggressive Behavior." *Journal of the American Statistical Association* 100 (471): 1054–76.

Wilkinson, Leland, and Task Force on Statistical Inference. 1999. "Statistical Methods in Psychology Journals: Guidelines and Explanations." *American Psychologist* 54 (8): 594–601.

Yuan, Ke-Hai, and Kentaro Hayashi. 2010. "Fitting Data to Model: Structural Equation Modeling Diagnosis Using Two Scatter Plots." *Psychological Methods* 15 (4): 335.