Parameterizing structural equation models as Bayesian multilevel regression models: An
example with the Global Multidimensional Poverty Index

James Ohisei Uanhoro

Department of Educational Studies, Ohio State University

Author Note

James Ohisei Uanhoro is a doctoral candidate in the Quantitative Research, Evaluation and Measurement program. Correspondence concerning this article should be addressed to James Ohisei Uanhoro (email: uanhoro.1@osu.edu), Center on Education and Training for Employment, 1900 Kenny Rd, Columbus, Ohio 43210. I thank Professor Ann O'Connell for providing feedback on an earlier draft of the manuscript. And I thank Professor Paul De Boeck, whose item-response theory course was helpful in coming to my current approach to a broad class of measurement models.

Abstract

The goal of this paper is to frame structural equation models (SEMs) as Bayesian multilevel regression models. Framing SEMs as Bayesian regression models provides an alternative approach to understanding SEMs that can improve model transparency and enhance innovation during modeling. For demonstration, we analyze six indicators of living standards data from 101 countries. The data are proportions and we develop confirmatory factor analysis as regression while accommodating the fact that the data are proportions. We also provide extensive guidance on prior specification, which is relevant for estimating complex regression models such as these. Finally, we run through regression equations for SEMs beyond the scope of the demonstration.

Parameterizing structural equation models as Bayesian multilevel regression models: An example with the Global Multidimensional Poverty Index

Introduction

Structural equation modeling (SEM) refers to a broad class of models that includes path analysis, confirmatory factor analysis, latent variable regression, and several latent variable models. A unifying theme for the different models that fall under the umbrella of SEM is that variance-covariance matrices of observed variables are often sufficient to perform analysis; hence, the alternative and more informative name: covariance structure modeling/analysis. In fact, for simpler structural equation models (SEMs), it is common to present SEM as a decomposition of a variance-covariance matrix of observed variables into a number of different matrices depending on the specific analysis being performed (Bollen, 1989).

In this paper, we show that basic SEMs can be parameterized and solved as (multilevel) (non-linear) regression models using the raw data.¹ This is immediately obvious for path analysis; the solution to a single path analysis model can be obtained from several univariate regression models. However, this equivalence is not as apparent for more complicated models like confirmatory factor analysis or latent variable regression models.

Before delving into our attempt at reparameterization, we lay out the rationale for this attempt. Regression analysis is the *workhorse of applied statistics*; this description of regression is not original (e.g. Ruppert, 1987; Smithson & Verkuilen, 2006). And most researchers with quantitative training are familiar with the regression model. Regression analysis is also one of the most studied statistical approaches. On the other hand, SEM is traditionally studied and developed within quantitative psychology, education and sociology. By unifying the approaches, researchers can learn from and contribute across their traditional areas of expertise.

Additionally, researchers trained in statistical modeling learn a language to

¹ Similarly to Gelman and Hill (2007), we use multilevel as an all inclusive term for mixed effects models whether they are nested or crossed in their data structure.

describe their models. This language to speak on models may arise from either *rote learning* with retention as an outcome or *meaningful learning* with understanding and the ability to transfer knowledge to new problems as outcomes (Mayer, 2002). One way to gain an understanding of models is to understand their mathematical representations. Given that the standard mathematical representation of SEMs requires knowledge of matrix algebra, a mathematical understanding of these models is beyond a substantial number of applied researchers. It often then is the case that students memorize certain bits of information to apply to substantive problems. And when faced with new problems, students find new information to memorize as guidance. On the other hand, connecting SEM to linear regression, i.e. connecting new knowledge to prior knowledge, offers a better chance at creating meaningful learning. Linear regression relies on a mathematical syntax that is comprehensible to quantitative applied researchers. This mathematical syntax, ($y = mx + c$ in its simplest form), allows for greater model transparency — compare different values of x and we can see expected differences in y . By embedding SEMs within this syntax, it is our hope that researchers gain additional clarity about SEMs. This greater transparency should lead to better understanding of models and their utility, and improve our ability to innovate when faced with new (and perhaps older) data analysis problems.

Additionally, instructors of quantitative methodology can introduce SEMs as precise specifications of a more general multilevel regression framework. And the reparameterization of SEMs as regression suggests easy ways to extend SEMs. Some extensions include a variety of distributional assumptions about observed and latent variables, and estimation of latent variables for complicated data structures, such as cross-classified data structures.

In this paper, we utilize Bayesian estimation as the computational approach for estimating SEMs as regressions. Although it is possible to estimate SEMs as regression using frequentist approaches, we prefer a Bayesian approach because of the wealth of information a Bayesian analysis yields relative to an analogous frequentist analysis (Kruschke, 2013). Additionally, commonplace general purpose Bayesian computation

engines (BUGS/JAGS/Stan) can handle most or all of the relevant regression problems for a multitude of data types (continuous/discrete) using their standard syntax. This is in contrast to frequentist approaches where there is need to identify/develop specific solutions for different types of problems; see further discussion in Box 1 below.

However, there are costs to the choice of a Bayesian approach. The first cost is computational resources, which manifests to the user as time. Bayesian methods for complicated models often require substantial computational resources and time. A second cost arises as a result of the use of Bayesian estimation: relative to traditional SEM estimation, there will be very few pre-packaged solutions for any given analysis, and the user will have to think through each part of model construction. However, thinking through each part of a model should improve the quality of the final model, and allows for innovation while modeling.

The Bayesian computational engine we use is Stan (Carpenter et al., 2017) via CmdStan (Stan Development Team, 2016) from within the R programming language. Both Stan and R are software that can be freely downloaded from the internet and are available on both Windows and Unix operating systems, e.g. Mac, Linux. A cost of using Stan (or other commonplace Bayesian computation engines) is that to estimate these models, the practitioner will have to wear the hat of a programmer. However, this is increasingly the way of applied statistics. To reduce the burden on practitioners, we provide code in the supplementary materials, and we have crafted the code such that readers can modify and apply the code to similar problems with other datasets.

We note that there is a history of methodologists connecting SEMs to multilevel regression. Rovine and Molenaar (2000) noted that one can apply the SEM approach to estimate a class of multilevel models known as growth curve models. And a recent book (Grimm, Ram, & Estabrook, 2016) covers growth curve models extensively, presenting both multilevel and SEM approaches. Curran (2003) extensively discussed the similarities between SEMs and multilevel models noting, “Any two-level linear multilevel model can be estimated as a structural equation model given that this is essentially a data management problem.” The generalized linear latent and mixed models framework

(GLLAMM, Rabe-Hesketh, Skrondal, & Pickles, 2004) in Stata heavily leverages the connection between both model classes. Additionally, the dominant school of Bayesian SEM documented in the text by Song and Lee (2012) explicitly develops SEMs as a series of regression with latent variables as predictors of item responses and or as outcomes depending on the specified model. The field of item-response theory (IRT, which has connections to confirmatory factor analysis, Kamata & Bauer, 2008; Takane & de Leeuw, 1987) also has the explanatory IRT approach (De Boeck & Wilson, 2004) which frames item-response modeling as generalized nonlinear mixed models.

What then is our contribution? We go beyond establishing the similarity between both models. We show how framing SEM as regression allows the modeler to think about SEM parameters differently. Additionally, we take advantage of Bayesian estimation to show how the modeler can move beyond standard model treatments to ask and answer new questions of data. We also focus extensively on prior specification in the models we run. We show how one can apply priors using a combination of subjective and objective knowledge of the data (prior to observing the data) while avoiding unprincipled biasing of model insights.

To accomplish the above, we analyze living standards data from the 2019 multidimensional poverty index (Alkire, Kanagaratnam, & Suppa, 2018). The data are the prevalence of deprivation (proportions) on six indicators of living standards for 101 countries. We state the broad questions we intend to answer using these data. Then we analyze the data from a regression standpoint, making connections to SEM during the analysis. The goal is to show how the regression framework matches the SEM model, while allowing for greater transparency and innovation during modeling. Importantly: the data are proportions and we show that despite achieving perfectly desirable model fit indices from the standard Gaussian confirmatory factor analysis (CFA), assuming the data are beta-distributed better accommodates the data and yields substantively different model insights. Given that the analytical approach is Bayesian, we provide guidelines for researchers for specifying priors – these guidelines are applicable to a wide variety of SEM models.

Beyond the analysis of the MPI, we provide a list of regression model equations that match basic SEMs followed by some discussion. It is our hope that the presentation of these basic models as regression models will provide researchers with an entry point to further explore this framing, and provide practitioners and students with new ways of understanding these models.

We expect that the reader who benefits the most from this paper is one who is (somewhat) familiar with Bayesian estimation, multilevel models and generalized linear models. Hence, this paper may be best for the individual who has to teach SEM to others, as the paper provides an alternative model conceptualization for SEMs with the goal of increased transparency. However, we take an expository approach to writing the paper. Hence the minimal requirement for this paper is a strong understanding of linear regression models.

Box 1: Why avoid frequentist approach to framing SEMs as regression models? Framing SEMs as multilevel models without Bayesian estimation is possible but can easily become a tedious task. To frame SEMs as multilevel models without Bayesian estimation, one needs to be able to estimate both nested and crossed random effects with regression equations that are non-linear in their parameters. Computational inference, such as bootstrapping, would be necessary for parameters like variance components or covariance matrices that take on a different relevance in an SEM-as-multilevel-model context. Such computational inference can increase program runtime, even making frequentist approaches more of a time-sink than a Bayesian approach. Adaptive quadrature might be required for estimation with non-normal data. The program would still have to account for a variety of error covariance structures. A variety of distributions would have to be programmed for different data types. Finally, different accommodations would be necessary for different data types – examples include options for handling zero-inflation and censoring with count data. One would have to program a computation engine, complete with the adequate choice of optimizers, or they would have to use a hodgepodge of statistical packages/tools depending on the problem at hand. On the other hand, commonplace

general purpose Bayesian computation engines (BUGS/JAGS/Stan) can handle most or all of these problems using their standard syntax.

Analyzing living standards using the Multidimensional Poverty Index

The multidimensional poverty index (MPI, Alkire et al., 2018, data retrieved from: <https://ophi.org.uk/multidimensional-poverty-index/data-tables-do-files/>) is an index that assesses 101 developing countries using 10-indicators of poverty. The MPI has three dimensions: health with two indicators, education with two indicators, and living standards with six indicators. We focus on the living standards indicators. The raw data for these indicators are the prevalence of deprivation depicted in Figure 1 with descriptive statistics in Table 1. As is clear from Figure 1, the data are not normally distributed.

Figure 1. Prevalence of deprivation for the six indicators of living standards.

We have three substantive questions of these data:

1. What is the ranking of the indicators in terms of prevalence, from least to most prevalent?

Table 1

Descriptive statistics for the prevalence of deprivation for the six living standard indicators in the 2019 Global Multidimensional Poverty Index

Indicator	n	Mean (SD)	Median (MAD)	Minimum	Maximum
Cooking fuel	100	52% (36)	57% (55)	0%	99.5%
Sanitation	101	42% (30)	42% (42)	1.3%	97.4%
Drinking water	101	27% (20)	25% (26)	0.3%	70.3%
Electricity	100	32% (33)	19% (28)	0%	95.6%
Housing	100	39% (30)	37% (38)	0.8%	98.2%
Assets	101	18% (19)	13% (17)	0%	70.1%

Note. MAD: Median absolute deviation. The n varies as one country was missing values for cooking fuel, electricity and housing.

2. Which indicators are most informative of a country's living standard, from least to most informative?
3. What are the thirty countries with the lowest living standards?

In addition to the substantive questions, we intend to elaborate on the following methodological/pedagogical points:

1. motivating a confirmatory factor analysis (CFA) from the ground-up via regression modeling;
2. the importance of aligning the CFA with the distribution of the data regardless of the model fit indices;
3. prior specification for model parameters;
4. using posterior distributions to answer substantive questions probabilistically.

We develop a series of models, with successive models building on their predecessors to answer the substantive questions and elaborate the methodological points above. After we present our final model for these data, we close this section with a discussion of the model insights: which insights follow entirely from the model and which insights require justification beyond what the data can provide?

A basic multilevel regression model

We begin our analysis by transforming the data from wide form to long form as shown in equation 1:

$$\begin{array}{|c|c|c|c|c|} \hline & \text{Country} & \text{Cook.} & \text{Sanit.} & \text{Drink.} \\ \hline \text{Armenia} & 2.7 & 23.5 & 2.1 & \\ \hline \text{Ukraine} & 5.1 & 2.3 & 2.5 & \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|c|} \hline \text{Country} & \text{Indicator} & \text{Prevalance} \\ \hline \text{Armenia} & \text{Cook.} & 2.7 \\ \hline \text{Armenia} & \text{Sanit.} & 23.5 \\ \hline \text{Armenia} & \text{Drink.} & 2.1 \\ \hline \text{Ukraine} & \text{Cook.} & 5.1 \\ \hline \text{Ukraine} & \text{Sanit.} & 2.3 \\ \hline \text{Ukraine} & \text{Drink.} & 2.5 \\ \hline \end{array} \quad (1)$$

A basic regression model for these data is: $p_{ic} = a + t_i + k_c + e_{ic}$; the prevalence of deprivation (p) for indicator i and country c (p_{ic}) is a function of an overall average a , the deviation of each indicator from the average (t_i), the deviation of each country from average (k_c) and the error term (e_{ic}). We adopt the following notation for this model:

$$p_{ic} \sim \mathcal{N}(a + t_i + k_c, \sigma) \quad (2)$$

where we assume the prevalences are normally distributed with a linear regression function $a + t_i + k_c$ and a residual standard deviation, σ . This is a standard regression problem. One possibility is to estimate the solution using five dummy codes for the standard of living indicators, and 100 dummy codes for the 101 countries.

However, we note that 100 is a large number of dummy codes, and the multilevel modeling framework offers a parsimonious approach to obtaining k_c , that we treat k_c as a *random effect*. We instead assume that $k_c \sim \mathcal{N}(0, \lambda)$; the (centered) prevalence for each country is normally distributed with a mean of zero, and a constant variance, λ^2 (see Box 2 for rationale for normality assumption). Instead of estimating 100 dummy coefficients or parameters, we estimate a single parameter for the constant variance.

Adopting typical multilevel language, k_c would be a *random-intercept* with a variance of λ^2 .

The updated model is: $p_{ic} \sim \mathcal{N}(a + t_i + k_c, \sigma)$, $k_c \sim \mathcal{N}(0, \lambda)$. We re-write this model to allow an alternative interpretation of the model:²

$$p_{ic} \sim \mathcal{N}(a + t_i + \lambda k'_c, \sigma), \quad k'_c \sim \mathcal{N}(0, 1) \quad (3)$$

Under this formulation, k'_c are the standardized averages for each country, and λ is then the relation (coefficient) between the standardized averages and the country's scores on any given indicator. Consider two countries, assuming one country has an average standardized score of 0, and the other has an average standardized score of 1, λ is the expected difference in prevalence on any given item. Equation 3 is akin to a one-way ANCOVA: t_i is categorical predictor (dummy variables), and k'_c is a continuous variable with coefficient λ . Recall that λ doubles as the standard deviation of the countries' average scores when they are unstandardized.

One can estimate the model in equation 3 in any multilevel regression software. And we have worked our way to this particular because it is identical to the *parallel model* within the CFA context.

Box 2: Why assume the means are normally distributed? There are multiple potential justifications for the normality assumption on each country's average scores. One based on central limit theorem is that if the country's average scores are the result of several minor additive features, their averages tend towards a normal distribution (e.g. Birnbaum, 2008, sec. 17.7). Another rationale for assuming normality is that if all we are willing to assume about the scores is that they are unbounded i.e. with range $(-\infty, \infty)$ and have finite variance, then the least surprising or *maximum entropy* distribution is the normal distribution (e.g. Frank, 2009; McElreath, 2020). If the practitioner has more information, then they may choose another distribution for the countries' average scores. In our case, the prevalences are between 0 and 100, so the maximum entropy justification does not suffice. Another justification would be that the

² This equivalence follows from the fact that $\text{Var}(a \times \mathbf{x}) = a^2$ when $\text{Var}(\mathbf{x}) = 1$.

normality assumption is simply a *device* for efficiently estimating the coefficients of a factor/grouping variable (Hodges & Clayton, 2011). Similarly to the previous justification, one may assume another distribution for the country's averages e.g., Student's t may be preferable if we assume that some of the countries are outliers such that they are unduly influential on the estimation of the country averages. Stan and Bayesian estimation in general permit alternative distributional assumptions about random intercepts. And one can use Bayesian model comparison methods (Piiironen & Vehtari, 2017) to compare models based on different distributional assumptions about random effects. Model assessment via residual diagnostics (O'Connell, Yeomans-Maldonado, & McCoach, 2015) or posterior predictive checking (Gelman, Goegebeur, Tuerlinckx, & Van Mechelen, 2000; Gelman, Meng, & Stern, 1996) can also be helpful in this regard. The assumption based on the central limit theorem is a staple of psychometric practice. However, the final assumption that normality is simply an efficient estimation device grants the normality assumption no special status, and normality is simply an analytical convenience (Hodges & Clayton, 2011).

Parallel and tau-equivalent unidimensional CFA model

So far, we have a random-intercept multilevel model with dummy variables for the indicators. This model is equivalent to a parallel CFA. Within a traditional CFA framework, the data remain in their wide form. We have six items *items* corresponding to the six indicators of living standards. The item *intercepts* correspond to the mean of the items ($a + t_i$). The *latent trait* or *construct* which would be named (inverse) living standards is simply the standardized random intercept which varies by country, k'_c . The factor loading which is the *relation between the standardized latent trait and the items* is λ . Recall that from a multilevel regression perspective, λ is the standard deviation of the random intercept k_c or the coefficient of the standardized random intercept, k'_c . Finally, in a parallel CFA, the error terms are assumed to have the same variance across all items, σ^2 , or the model is homoskedastic from a regression perspective.

The above paragraph sheds some transparency on SEM language: latent variables

are random effects, or averages which have been estimated in a parsimonious manner. And loadings are a scaling factor for random effects. Item intercepts are averages which are often estimated as fixed effects. Any additional claims about the data, such as the naming of latent variables or aligning latent variables to constructs is never justified on the basis of the model.

Given the data in Figure 1, the different standard of living indicators clearly have different variances, hence it is wise to update our model to permit heteroskedasticity:

$$p_{ic} \sim \mathcal{N}(a + t_i + \lambda k'_c, \sigma_i), \quad k'_c \sim \mathcal{N}(0, 1) \quad (4)$$

The error variance now differs by indicator. This model is known as the tau-equivalent model in CFA (Brown, 2006, p. 239) – the items have the same relation to the factor (λ), however the items differ in the means and error variances.

Figure 2. Relationship between country averages and prevalence of deprivation for each indicator based on the tau-equivalent model. Each spike on the x-axis represents the standardized average for a country. For each indicator, the relation between country averages and prevalence is the same, hence the parallel lines. Given the earlier stated relation between this model and ANCOVA, this pattern is the parallel slopes assumption of ANCOVA (Glass, Peckham, & Sanders, 1972).

Data analysis. To demonstrate the equivalence of the parallel and tau-equivalent CFA models to multilevel regression, we tested the both CFA models using lavaan (Rosseel, 2012) (SEM software in R), and both multilevel models above using glmmTMB (Brooks et al., 2017), multilevel software also in R. Prior to analysis, we rescaled the prevalences from 0 – 100 to 0 – 1. In lavaan, we set the missing data treatment to full-information maximum likelihood; this is equivalent to using all available rows in the long-form data for multilevel regression. Additionally, the multilevel model has to use maximum likelihood (ML) estimation not restricted maximum likelihood (REML) for the results to match standard SEM results.

The log-likelihood of the parallel and tau-equivalent models were 165.8533 and 181.1196 respectively in both software. The loading returned from lavaan and the random effect standard deviation from glmmTMB was 0.2449 (parallel model) and 0.2196 (tau-equivalent model). The error variance from the parallel models was 0.02077 in both software. The error variances for the tau-equivalent models varied by item and were identical across software to three decimal places. The item means and their standard errors were also identical across software to four decimal places.

Next, we visualize the relation between k'_c and the countries' scores from the tau-equivalent model in Figure 2. As evident from the graph, we see that countries who are a point different in their standardized average scores are about 20% points apart in their prevalence, corresponding to $\lambda = 0.2196$. Based on Figure 2, the cooking indicator had the highest prevalence on average, while the asset indicator had the lowest prevalence on average.

Model assessment. Without resorting to model diagnostics, we know that this model is grossly misspecified in a number of ways. From Figure 1, the data are clearly not normally distributed. This has implications for interpreting model results. For example, the expected prevalence of asset deprivation is negative for countries with very low average scores – this is impossible. Moreover, non-normality may have implications for inference.

From an SEM perspective, the model-data fit was poor with SRMR above 0.25,

RMSEA above 0.3, CFI and TLI under 0.8, a χ^2 p-value under .001 rejecting the null hypothesis of model-data fit. Within an SEM framework, we can begin to rectify this model by releasing certain constraints on the model to improve model-data fit, e.g. permit the loading to vary by item. Additionally, for non-normal data, there are several robust ML techniques (Savalei, 2014) and bootstrap methods (Falk, 2018) to help with adequate inference. However, these robust approaches for non-normality do not improve the model-data fit, they simply permit adequate inference under non-normality.

From a regression perspective, the big problem is the distribution of the data: the data are proportions bounded at 0 and 1. From Figure 1, the data are sometimes close to the boundaries, hence, a normal fit to the data will be inadequate. For data such as these (bounded with notable *ceiling* and *floor effects*), a more adequate distribution would be the beta distribution (Smithson & Verkuilen, 2006, see Box 3 below). Once we analyze the data using the correct distributional assumptions, we can revisit extensions to the basic model above.

Box 3: Why the beta distribution? The beta distribution is a distribution for continuous data with range of (0, 1) such as proportions. Other applications of the beta distribution include grade point average, averaged Likert data – the data would first be re-scaled to the 0 – 1 scale. Same as the normal distribution is defined by its mean and variance, the beta distribution can be defined by its mean and *sample size* (Kruschke, 2011). The mean also known as the proportion is the mean of the data on the 0 – 1 scale. The sample size (> 0) relates to how certain one is about this mean, hence the alternative name: *precision*. As seen in Figure 3, the higher the sample size, the less dispersed the data. And when the sample size is less than 1, the data can take on a U-shaped, matching the distribution of the cooking fuel prevalence. These properties make the beta distribution a highly flexible distribution for modeling bounded continuous data.

Figure 3. Visualizing the beta distribution with different parameters, $\text{Beta}(\text{mean} = x_1, \text{sample size} = x_2)$. The vertical dashed lines is the mean. As shown in this graph, the distribution can take on a variety of shapes depending on value of both parameters.

A beta distributed tau-equivalent CFA

Our working model so far is a simple heteroskedastic random-intercept regression model, and we may be tempted to simply update the model to use the beta distribution in the form below:

$$p_{ic} \sim \text{Beta}(a + t_i + \lambda k'_c, n_i), \quad k'_c \sim \mathcal{N}(0, 1) \quad (5)$$

where the mean is as before and the sample size varies by indicator to account for the fact that the different indicators are differently dispersed.³ If we analyzed these data as above in equation 5, we run into a major problem. The mean of the beta distribution must fall in the 0 – 1 range. However, there is currently no guarantee that the mean above will lie in this range. In fact, as seen in Figure 2, the regression

³ The beta distribution is still able to capture heteroskedasticity even if we do not vary the sample size parameter as the variance depends on both the mean and sample size. However, explicitly modeling the sample size allows us capture substantial differences in the shapes of the different variables (Smithson & Verkuilen, 2006).

prediction was sometimes less than zero or greater than one. The standard solution (from generalized linear modeling) is to calculate the mean on an unconstrained scale, then transform the unconstrained mean to lie between 0 and 1. To this end, we use the logit-link $\left(\exp(.) / (1 + \exp(.))\right)$ as is standard in beta regression (Ferrari & Cribari-Neto, 2004; Smithson & Verkuilen, 2006). Our updated model is now:

$$p_{ic} \sim \text{Beta} \left(\frac{\exp(\eta_{ic})}{1 + \exp(\eta_{ic})}, n_i \right), \quad \eta_{ic} = a + t_i + \lambda k'_c, \quad k'_c \sim \mathcal{N}(0, 1) \quad (6)$$

This mean on the unconstrained scale (η_{ic}) can take on any real number, while the logit link guarantees that the mean passed to the beta distribution will lie between 0 and 1. This model is akin to a heteroskedastic multilevel linear regression model in equation 4, with the exception that we have modified the distribution to match the data, and we now have a logit link function to ensure the mean of the data lies between 0 and 1.

One minor adjustment is necessary prior to running this model. The beta-distribution can only fit data between 0 and 1, exclusive. From the descriptive statistics in Table 1, there is at least one country with 0% cooking fuel, electricity and asset deprivation. For each of these indicators, we rescaled the data using the formula: $(y_c \times (n - 1) + 0.5) / n$, where y_c is the prevalence in country c and n is the number of countries for which data are available on a given indicator. This re-scaling recommended by (Smithson & Verkuilen, 2006) barely affects the data beyond ensuring that all the data points lie between 0 and 1.

Data analysis. We analyzed the multilevel beta regression in equation 6 using glmmTMB. For comparison, we re-ran the heteroskedastic multilevel linear regression (equivalent to tau-equivalent model in lavaan) on the same outcome (with all values transformed to fall between 0 and 1) as the beta regression model in glmmTMB. The AIC and BIC for the normal distribution model were -343 and -285 respectively. For the beta-distribution model, the AIC and BIC were -886 and -828 respectively. This is a huge improvement in model fit for models with the same number of parameters. By changing the distribution, the degree of misspecification has dropped drastically.

Figure 4. Relationship between country averages and prevalence of deprivation for each indicator based on the beta-distributed tau-equivalent model. For each indicator, the relation between country averages and prevalence is the same on the unconstrained scale (right panel), but varies on the proportion scale (left panel).

Moreover, the predictions from the multilevel beta regression never go below 0 or exceed 1 as seen in the left panel of Figure 4.

Although, the relative ordering of the indicators in terms of prevalence is the same with the earlier models, the expected prevalence predicted by the beta model does not always match the expected prevalence from the models assuming normality. Given our use of the logit link, one can calculate the average prevalence by indicator for the beta model using: $\exp(a + t_i) / (1 + \exp(a + t_i))$. The average prevalence of deprivation was: cooking fuel (52% from normal model, 51% from beta model), sanitation (42%, 39%), drinking water (27%, 20%), electricity (32%, 26%), housing (39%, 36%) and assets (18%, 12%). Notable differences in average prevalence predicted by the models exist for drinking water (7% point difference), electricity (6%) and assets (6%); these happen to be the skewed indicators. Comparison of the average prevalence to the descriptive statistics in Table 1 shows that the expected prevalence returned by the beta model is closer to the median when the indicator is skewed, while the normal distribution model

returns the mean. The median is known to be more robust than the mean as an estimator of the location (average) when data are skewed, hence the expected prevalence based on the beta distribution are relatively robust to the skew in the data.

Next steps. From a standard (linear) multilevel regression perspective, we would need additional predictors to further improve the heteroskedastic multilevel beta regression in equation 6. However, from an SEM perspective, we can further improve our model by assuming that the relation between the standardized averages and the expected responses varies by indicator. This model is known as a congeneric CFA model, and we cover it in the next section.

Bayesian multilevel equivalent of a congeneric CFA

The most common CFA for items assumed to load onto a single construct is the congeneric CFA. In this model, the items have different means, different variances, and different relations with the latent trait:

$$p_{ic} \sim \mathcal{N}(a + t_i + \lambda_i k'_c, \sigma_i) , \quad k'_c \sim \mathcal{N}(0, 1) \quad (7)$$

This model differs from all other models in a fundamental way: the model is non-linear in its parameters, i.e we need to multiply λ_i by k'_c .⁴ Hence, this model and models that allow the relation between the averages and indicators to vary by each indicator cannot be fit with most multilevel regression software, as most multilevel regression software are linear multilevel regression software. Bayesian modeling can handle non-linear equations and we develop the model in equation 7 above within a Bayesian framework.

Prior specification. The first task is to determine the priors for the parameters of interest. We can think of priors as plausible values for parameters of interest (McElreath, 2020), and it is common to specify priors using statistical distributions. To specify the priors, we focus on the variances related to different parameters. According

⁴ Although we presented the earlier models with a product of parameters: $\lambda k'_c$, note that multilevel regression software estimates unstandardized random effects (k_c) with λ as its standard deviation.

to equation 7, the variance in the outcome is due to three sources:

1. the differences between items, t_i : If all the items are similarly prevalent, this variance would account for a relatively small amount of variance in the overall response. We term this variance parameter, σ_t^2 .
2. the differences between individuals captured in the model as $\lambda_i k'_c$ with a variance that varies by item, λ_i^2 ; and
3. the differences due to factors unaccounted for by the model with variance that varies by item, σ_i^2 .

We now provide some heuristics to specify plausible values for each of these variance components. Given our assumption that the data are bounded, we can assume that the item means are spread equidistantly across the range of the data implying that the item means follow a discrete uniform distribution. The variance of a discrete uniform variable is $\left[(b - a + 1)^2 - 1\right]/12$, where a and b are the lower and upper limits (bounds on the data) respectively. Hence, the variance of the item means is then 0.25, $\left[(1 - 0 + 1)^2 - 1\right]/12$. In reality, we should expect the item means to have less variance than 0.25 since item means do not tend to be equidistantly spread across the range of the response scale. Hence, we assume 0.25 to be an upper limit on the variance of the item means, i.e. $0 < \sigma_t \leq 0.5$ ($\sqrt{0.25}$).

We again work with the range of the data to determine the prior for σ_i^2 . Given that we assume that individual responses about each item (ϵ_{ic}) are normally distributed, we can invoke the 68 – 95 – 99.7 rule: for normally distributed data, about 95% of observations will lie within 1.96 standard deviations of the mean. Using this heuristic, assuming a centrally placed item with a mean of 0.5, 0.255 ($.5/1.96$) is a prudent choice of standard deviation of responses about the item mean if 95% of the responses are to fall in the $[0, 1]$ range. The variation in these responses is due to variability in both individuals ($\lambda_i k'_c$) and unmeasured factors. All observations lie within 0 and 1 for our data suggesting a smaller standard deviation than 0.255. Moreover, item means will not be 0.5 suggesting that 0.255 is a high estimate for the

standard deviation of the responses about each item mean. Hence, we will treat 0.255 as the upper limit on σ_i , $0 < \sigma_i \leq 0.255$. This is a limit that is only likely to be reached if all items have a mean of 0.5, and the variability about the responses due to their average scores is 0 (or $\lambda_i = 0$).

For the last component, since k'_c is a standardized variable, we can expect that 95% of values will lie between -2 and 2 . Given that λ_i are coefficients for k'_c and there is a $0 - 1$ range for the data, we can divide the range of the data ($1 - 0 = 1$) by the expected range of k'_c ($2 - -2 = 4$) to obtain a maximum limit for λ_i . This suggests a maximum limit of 0.25 ($1/4$) for λ_i .

At this point, we have obtained plausible ranges for the variance of the item means ($0 < \sigma_t \leq 0.5$), the variance due to individual differences ($0 < \lambda_i \leq 0.25$), and the variance due to unmeasured factors, $0 < \sigma_i \leq 0.255$. Note that the three preceding inequalities are presented using standard deviations. The procedures we have undertaken to arrive at these estimates rely on information about the response scale of the data, and assumptions we have made about the patterns in the data. Where more information is available about the data, one may make even stronger assumptions about these parameters. At this point, we define a Bayesian model for the data:

$$\begin{aligned}
 \text{Likelihood: } p_{ic} &\sim \mathcal{N}(a + t_i + \lambda_i k'_c, \sigma_i) \\
 \text{Priors: } a &\sim \mathcal{N}(0, 1), \quad t_i \sim \mathcal{N}(0, \sigma_t), \quad \sigma_i \sim \mathcal{N}^+(0, 0.13), \quad k'_c \sim \mathcal{N}(0, 1) \\
 \lambda_i &\sim \text{log-Normal}(\text{median} = 0.1, \text{scale} = \ln(0.25/0.1)/1.645) \\
 \text{Hyperpriors: } \sigma_t &\sim \mathcal{N}^+(0, 0.255)
 \end{aligned} \tag{8}$$

The likelihood line of equation 8 specifies our claims about the distribution of data. The next line specifies the prior distributions, or our a-priori beliefs about the model parameters.

The half-normal ($\mathcal{N}^+$) prior on σ_i ensures σ_i will be positive. The choice of 0.13 ($0.255/1.96$) as the standard deviation of this prior is based on the belief that $\sigma_i \leq 0.255$. Recall that for the normal distribution, about 95% of observations are

within 1.96 standard deviations of the mean. By selecting 0.13 as the standard deviation of the half-normal prior, we are assuming a 95% chance that σ_i is a positive value less than 0.255. This assumption may be wrong, so we leave a 5% chance that σ_i is greater than 0.255. This practice of using a 95% plausible range for parameters to arrive at a prior distribution is recommended by Greenland and Mansournia (2015).

This overall mean (a) has a normal prior with a wide variance (1) on it, thus permitting a wide range of values for the overall mean. The prior on the standard deviation of item means (σ_t) is half-normal with a mean of zero and standard deviation, 0.255. This choice is based on the same considerations for standard deviation of the prior on σ_i . Since σ_t should be less than 0.5, setting the standard deviation of its prior to 0.255 ($0.5/1.96$) assigns a 95% chance that σ_t is less than 0.5.

Figure 5. Visualizing log-Normal (median = 0.1, scale = $\ln(2.5)/1.645$). The plot on the left shows the shape of the distribution. The plot on the right shows what proportion of the distribution is less than potential values of λ_i .

The standardized average scores, k'_c , are assumed standard-normal as with standardized variables. The prior on λ_i is assumed log-normal and deserves special consideration. Parameters such as λ_i can be difficult to estimate especially if the data are uninformative – on average, this is more likely to occur at small sample sizes. The log-normal distribution constrains λ_i to be positive and can also keep the estimate away from exactly 0, a reasonable assumption if we assume all the indicators positively relate to the average scores. This assumption is highly reasonable in our example, as one would expect the different indicators to be positively related to each other. One

approach for choosing parameters for the log-normal distribution is to postulate the median and an upper limit for the parameter. In our case, we have assumed a median value of 0.1 for the λ_i , i.e. a 50% chance that λ_i exceed or are less than 0.1. This implies that the shift from -2 to 2 in k'_c relates to a 40% point change $((2 - -2) \times .1)$ in the prevalence scores. Given that we have specified an upper limit of 0.25 for λ_i , we follow our previous approach and assume a 95% chance that λ_i are below 0.25. To achieve this, we set the log-normal scale parameter to $\ln(\text{upper limit}/\text{median})/1.645$.⁵ This results in a scale parameter of $\ln(0.25/.1)/1.645 = \ln(2.5)/1.645$. Setting the median and scale parameter in this way assigns a 50% chance that λ_i exceeds (is less than) 0.1, and a 95% chance that λ_i is less than 0.25. This choice has the additional effect of keeping λ_i away from zero (see Figure 5), a feature that is desirable for complex models such as this. As with other prior specifications, we have left some chance that we are wrong.

Estimates with 95% interval

Figure 6. Parameter estimates and 95% interval. The loadings from the Bayesian approach are consistently lower than the CFA loadings. This is due to the prior on the Bayesian loadings which pulls them towards the median of 0.1. This shrinking of the loadings also has the effect of narrowing the uncertainty intervals about the estimates.

⁵ 1.645 is to the 95% percentile of the standard normal distribution. For example, if we replaced 1.645 with 1.96 (the 97.5th percentile), we would be assuming a 97.5% chance that λ_i are less than 0.25.

Data analysis. The Stan syntax to estimate the models is available in the supplementary materials section. For computation, we drew 4000 posterior samples across 8 parallel chains, and retained 3000 samples within each chain for inference. This left 24,000 samples for inference for each estimated or computed parameter. On a laptop with a 3.1 GHz i7 processor running on Ubuntu, Stan converged in about 3 seconds. CmdStan performs model diagnostics using checking the potential scale reduction factor ($\hat{R}$), effective sample size (n_{eff}) and additional diagnostics; these diagnostics were satisfactory. We also inspected the rank plots (an improvement over traceplots, Vehtari, Gelman, Simpson, Carpenter, & Burkner, 2020, reported in the appendix) for the estimated parameters as an additional check of sampling convergence, and the rank plots were adequate.

We present parameter estimates and quantile intervals in Figure 6. We also present results for the standard congeneric model from lavaan for comparison. For item means and standard deviations, the parameter estimates were similar across models. For the relation between the average prevalence and the indicators which we have permitted to vary by indicator, the Bayesian estimates are lower than the estimates from lavaan, this is likely because the log-normal prior constrained the Bayesian estimates. However, constraining estimates in this manner is desirable for estimation efficiency (Davis-Stober, Dana, & Rouder, 2018) as suggested by the narrower Bayesian intervals.

Figure 7 shows the substantive implications of permitting the relation between country averages and indicators to vary by indicator. We see that countries show very large differences on indicators such as cooking fuel (#1) and electricity (#4) depending on their average scores; corresponding to the large loadings in Figure 6. At the other end, even countries with very high averages only have about 50 – 60% prevalence of drinking water (#3) and asset (#6) deprivation.

However, this model is again deficient as the data are clearly non-normal. Under zero and above 100% expected prevalences from the model in Figure 7 are impossible. Additionally, any inference based on this model will likely be inadequate. Within lavaan, one can improve inference by using a robust estimator or the bootstrap. Given

Figure 7. Relationship between country averages and prevalence of deprivation for each indicator based on the congeneric model (Bayesian regression). The relation between country averages and prevalence differs by indicator. Given the earlier stated relation between the tau-equivalent model and ANCOVA, this pattern is akin to the relaxation of the parallel slopes assumption of ANCOVA (Glass et al., 1972). Stronger slopes are bolder.

our Bayesian framework, we now present a beta congeneric CFA.

A beta distributed congeneric CFA

We present two potential options for a beta congeneric CFA. One is a beta regression analog of the non-linear and heteroskedastic multilevel regression model in equation 7:

$$p_{ic} \sim \text{Beta} \left(\frac{\exp(\eta_{ic})}{1 + \exp(\eta_{ic})}, n_i \right), \quad \eta_{ic} = a + t_i + \lambda_i k'_c, \quad k'_c \sim \mathcal{N}(0, 1) \quad (9)$$

The only change to the beta distributed tau-equivalent model in equation 6 is λ is allowed to vary by indicator.

The second approach utilizes observation-level random effects (OLRE, Harrison,

2014; Hinde, 1982) to capture indicator-specific variation:

$$\begin{aligned} p_{ic} &\sim \text{Beta}\left(\frac{\exp(\eta_{ic})}{1 + \exp(\eta_{ic})}, n\right), & \eta_{ic} &= a + t_i + \lambda_i k'_c + \sigma_i e_{ic}, \\ k'_c &\sim \mathcal{N}(0, 1), & e_{ic} &\sim t(\nu, 0, 1) \end{aligned} \quad (10)$$

where ν is the degrees of freedom of the t -distribution. The key differences from the first approach are the sample size parameter is now constant, and we introduced an observation-level error term (e_{ic}) with a scale parameter (σ_i) that varies by indicator. This is akin to a linear regression model where we estimate the error term as part of the model, this is common strategy with generalized linear models to capture case-level variation in the outcome (Harrison, 2014; Hinde, 1982). This option would not be possible under a standard covariance-matrix based approach to SEM, as one is unable to include predictors that vary at the level of both country and indicator. The main advantage of this approach over the first approach is that σ_i (the scale parameter of the t distribution) functions like a residual standard deviation parameter permitted to vary by item, with larger values connoting more variation in the prevalence of an indicator. Additionally, by setting e_{ic} to be t -distributed rather than normally distributed, we can account for outlying data points (West, 1984), as small values of ν (< 30) indicate the presence of outliers. Given these advantages, we focus on the second approach using OLREs going forward.⁶

Prior specification. In the normal distribution model above, we used information about the scale of the outcome to inform our choices of priors. However, all the parameters in beta distribution model above, except for the sample size (n_i), are on the logit scale. This makes specifying priors for these parameters a little more difficult as we cannot directly use information about the scale of the raw data.

Unless the item means are known to take on extreme values, a reasonable guess is that the average prevalence for each indicator falls between 5% and 95%. On the logit-scale, this is a range of $-3 \left[\ln(.05/(1 - .05)) \right]$ to $3 \left[\ln(.95/(1 - .95)) \right]$. We can

⁶ We performed both approaches and the second approach was superior as shown in the Appendix.

assume that the items are equidistantly spread within this range, thus the item means follow a discrete uniform distribution. The variance of the item means is then 4, $[(3 - -3 + 1)^2 - 1]/12$. In reality, the items means are unlikely to be spread as widely as this. In fact, we know that the average prevalence in our example are more likely to fall between 5% and 70% given the data in Figure 1. So 4 as variance estimate can serve as an upper estimate of the variance of the item means on the logit scale, $0 < \sigma_t \leq 2$.

Figure 8. Relation between loading and expected prevalence as a function of indicator prevalence and standardized average on logit-scale. The expected prevalence for countries with extreme standardized averages is extreme when the loading is 4 or 5.

Setting a prior on λ is more involved because of the logit-link function. We can use Figure 8 to guide our selection. In this figure, we consider two hypothetical countries with extreme standardized average scores of -2 and 2 . And we vary the indicator prevalence from 5% to 50% (varying from 50% to 95% simply produces the opposite of 5% to 50%). We also vary λ to see how the expected prevalence for both countries change. We find that even for an indicator with low prevalence (5%), when $\lambda = 4$, a country with a high standardized average will have near 100% chance of deprivation on

the indicator. Based on this example and other scenarios in Figure 8, λ need not exceed 4 and we set 4 as the upper limit on λ . However, should we have indicators with more extreme average prevalence (e.g. $< 5\%$), then the upper limit on λ may need to be higher. Alternatively, should we expect that the prevalences occupy a narrower range (e.g. most values fall between 20% and 80%), λ can be much lower than 4.

The full Bayesian model is then:

$$\begin{aligned}
\text{Likelihood: } p_{ic} &\sim \text{Beta}\left(\frac{\exp(\eta_{ic})}{1 + \exp(\eta_{ic})}, n\right), \quad \eta_{ic} = a + t_i + \lambda_i k'_c + \sigma_i e_{ic} \\
\text{Priors: } a &\sim \mathcal{N}(0, 2.55), \quad t_i \sim \mathcal{N}(0, \sigma_t), \quad k'_c \sim \mathcal{N}(0, 1) \\
e_{ic} &\sim t(\nu, 0, 1), \quad \sigma_i \sim t^+(3, 0, 1) \\
\lambda_i &\sim \text{log-Normal}(\text{median} = 0.5, \text{scale} = \ln(4/0.5)/1.645) \\
n &\sim \text{Gamma}(\text{shape} = 2, \text{rate} = 0.1) \\
\text{Hyperpriors: } \sigma_t &\sim \mathcal{N}^+(0, 1.02), \quad \nu \sim \text{Gamma}(\text{shape} = 2, \text{rate} = 0.1)
\end{aligned} \tag{11}$$

We placed a wide prior on the overall average as this is an easy parameter to estimate. The prior on the standard deviation of the item means σ_t is half-normal and assumes a 95% chance that σ_t will not exceed 2. The standardized averages are standard normal. We conservatively assumed a median value of 0.5 for λ_i . Given that we have a limit of 4 for λ_i , we can assign a 95% chance that $\lambda_i \leq 4$ by setting the scale to $(\ln(4/0.5)/1.645)$ as in the previous model. Finally, one thing we know about the sample size and degrees of freedom parameters is that they have to be positive numbers. The $\text{Gamma}(2, 0.1)$ prior allows a wide range of positive values for both parameters while keeping estimates away from zero (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). And the scale parameters (σ_i) have a half- t distribution which pulls them towards zero while simultaneously permitting these values to be large.

Data analysis. The Stan syntax to estimate the models is available in the supplementary materials section. We drew the same numbers of samples on the same computer. Stan converged in about 3 minutes. The model diagnostics were satisfactory.

Figure 9. Parameter estimates and 95% interval for Bayesian beta congenetic CFA (with OLRE) and normal congenetic CFA (with bootstrapped intervals). The expected prevalences and ranking of loadings and scale parameters differs across models.

We also inspected the rank plots (reported in the appendix) for the estimated parameters as an additional check of sampling convergence, and the rank plots were adequate. Additionally, the median of the degrees of freedom parameter for the OLREs (ν) was 18 suggesting that there were outliers in the data.

We present parameter estimates and quantile intervals in Figure 9. We also present results from a bootstrapped congenetic model in lavaan for comparison. The bootstrapped lavaan model returns the same parameter estimates as the earlier congenetic model fitted in lavaan, however, inference based on the bootstrapped model should be robust to the non-normality of the data. Additionally, the lavaan model had excellent global fit statistics: $\chi^2(9) = 12.4$, $p = .19$, SRMR of 0.014, RMSEA of 0.061

($P(\text{RMSEA} \leq 0.05) = .36$) and CFI and TLI above 0.99.

Comparing both models, there were differences in prevalence across models with the greatest differences occurring for electricity, indicator #4 and assets (#6). The prevalence of electricity deprivation (17% beta, 31% bootstrapped normal) and asset deprivation (9.2% beta, 18% bootstrapped normal) were such different that their 95% intervals did not overlap. The beta model estimates for electricity and asset deprivation were close to the median (19% and 13% respectively, see Table 1), and the median is preferable to the mean as a measure of location for skewed data such as these.

Note that separately from a comparison of raw estimates, the ranking of indicators by prevalence differs across both approaches. Although both models predict that asset deprivation (#6) had the lowest prevalence, the normal model suggests the next lowest indicator was drinking water (#3), while the beta model suggests it was electricity (#4). Hence, our substantive conclusions about the indicator prevalence (raw or ranked) changes depending on the model.

One is unable to compare the loadings from both models directly as they are on different scales. However, the ranking of the indicators by relation to the averages slightly differs across the models. The beta model suggests housing (#5) and asset deprivation (#6) have similar relation to the country averages as seen in the top left panel of Figure 9, while the normal model suggests housing deprivation (#5) has the stronger relation.

Also of interest are the scale parameters reported by the beta model. Cooking fuel had the largest value suggesting substantial variation in cooking fuel deprivation is unrelated to the country averages. At the other end, electricity had the smallest scale parameter suggesting that relative to other indicators, much of the variation in electricity deprivation is related to the country averages.

Additionally, having computed both loadings and scale parameters, one can “standardize” the loadings relative to the total variation in each indicator using the formula: $\sqrt{\lambda_i^2 / (\lambda_i^2 + \sigma_i^2 \nu / (\nu - 2))}$. Note that σ_i^2 need to be re-scaled by the degrees of freedom (ν) of the t distribution to obtain the residual variance.

Figure 10. Density of prevalence simulated from the beta congeneric CFA (with OLRE, 100 simulations) overlaid on density of original data (solid black line).

Model assessment. From a regression perspective, the beta model is clearly superior to the normal model (with or without bootstrapping). This is because the beta model accommodates the nature of the data. Moreover, we have seen that substantive conclusions about the data differ depending on the selected model, hence the differences between both modeling options is not purely of esoteric interests.

We conducted posterior predictive checking (Gelman et al., 2000, 1996) by overlaying data simulated from the model against the original data. The beta model performed much better than the normal model in this regard. Figure 10 shows how well simulated prevalences from the beta model matched the original prevalences, see comparison Figure for Gaussian model in Appendix. The results suggest that the beta model was able to capture the general shape of each indicator.

Addressing the substantive questions

We have built up to the beta-distributed congeneric model for these data. We began with standard multilevel regression modeling practices (including

heteroskedasticity), accounted for the non-normal nature of the data (beta distribution), and infused the concept of varying loadings (congeneric) from the SEM literature. We also included observation-level random effects to model case-level variation in the data. And posterior predictive checking suggested that the final model was able to capture the general shape of each indicator.

We now return to characterizing the data using the model. We restate our substantive questions:

1. What is the ranking of the indicators in terms of average prevalence?
2. Which indicators are most informative of a country's living standards?
3. What are the thirty countries with the lowest living standards?

We devote a section to each of these questions.

Figure 11. Top panel: Probabilistic ranking of indicators by prevalence. For example, there was a 94.4% chance that electricity deprivation had the second lowest prevalence. Bottom panel: Probabilistic ranking of indicators by strength of relation to living standards. For example, there was a 100% chance that drinking water deprivation had the weakest relation to living standards.

Ranking the indicators by prevalence. The beta congeneric model suggests the following ordering of indicator prevalences from lowest to highest: assets (9.2% deprivation), electricity (17%), drinking water (20%), housing (33%), sanitation (36%) and cooking fuel (51%). To obtain some certainty about these rankings, we examined each posterior iteration to identify the number of times the ranking within each iteration matched this average ranking. We present the results in the top panel of Figure 11. We see that the indicator with the greatest prevalence is unambiguously cooking fuel deprivation, while the indicator with the lowest prevalence is unambiguously asset deprivation. The greatest ambiguity in ranking occurs for housing and sanitation deprivation, however, the model suggests that sanitation deprivation is the more common phenomenon.

Figure 12. Relationship between country averages and prevalence of deprivation for each indicator based on the beta congeneric model (Bayesian regression). The relation between country averages and prevalence differs by indicator on both the unconstrained (not depicted) and probability scales. Stronger relations are bolder.

Ranking the indicators by relation to living standards. We can think of the standardized averages as a proxy for living standards. However, higher indicator prevalences connote deprivation, so they are negatively related to living standards.

Hence, we treat living standards as standardized averages multiplied by -1 . We plot the relation between living standards and each indicator in Figure 12. We see that countries with relatively lower living standards have very high electricity (#4) and cooking fuel (#1) deprivation, while countries with higher living standards have much lower electricity and cooking fuel deprivation. The relation of indicators with living standards is weaker for other indicators.

We examined each posterior iteration to determine the ranking of each indicator relation to the standardized averages across iterations. We present the results in the bottom panel of Figure 11. There was a 65% chance that electricity deprivation had the strongest relation to living standards. This suggests that when examining countries in this dataset, differences in electricity access was most informative about differences in living standards across countries. The next strongest indicator was cooking fuel, and the weakest indicator was unambiguously drinking water. And there was some ambiguity about the ordering of asset, sanitation and housing deprivation.

Identifying the 30 countries with the lowest living standards. We briefly lay out the procedure for answering this question:

- For each iteration in the 24,000 posterior iterations: Code every country that has a top 30 value of k'_c as 1, otherwise code the country as 0.
- Across the 24,000 iterations, average the binary indicator for each country to obtain the country's probability of a top 30 score.

We present the results of this procedure in Figure 13. 18 countries had more than 99% chance of a bottom 30 living standard, and there were 21 countries with more than 95% chance. These probabilities suggest very high certainty for the majority of our selection.

Of the 30 countries, there were four (IDs = 81, 76, 95, 83) with probability of a bottom 30 living standard between 50% and 80%. And there was an ambiguous country (ID 82) with more than 50% chance of a bottom 30 score but its living standard was not among the 30 lowest scores. Although the data suggest a ranking of these five

Figure 13. Probability each country had a bottom 30 standard of living. Of the 30 countries with the highest probabilities, 24 had above 95% chance of having a bottom 30 score indicating high certainty about the selection of countries. Countries with under 80% and above 50% chance are labelled using their exact probabilities on the left and their row IDs in the dataset on the right.

countries, it would be wise to rely on information outside the model in distinguishing these countries when making a selection. For example, a comparison of posterior samples suggested that there was only a 53% chance that country with ID 83 had a lower living standard than country with ID 81.

Brief discussion of analysis

One should be skeptical of certain claims in the preceding analyses. The most important claim to be skeptical of is that the country averages are (a proxy for) living standards. We only know that we have estimated the country averages across the different indicators. The claim that these country averages are (a proxy for) living standards neither arises from nor is supported by the data analysis. This important point is one researchers often take for granted, assuming validation via CFA (and related procedures) renders meaning to (standardized) averages, when a clear understanding of the regression models above shows that validation via CFA does no

such thing. The “successful validation” of the meaningless theory of gavagai by (Maul, 2017) demonstrates this point using similar survey validation techniques.

Accordingly, the next claim to be skeptical of is that we have identified the relation of the different indicators to living standards. This claim rests on the earlier claim that the averages are a proxy for living standards. All we know based on the analysis is the relation between the country averages and the individual indicators of living standards. Human judgement may lead one to prioritize one indicator over others regardless of the model findings. Recall that this model does not account for the consequences of the different indicators.

One *robust* claim based on this model is the ranking of the indicators by prevalence of deprivation. This has clear human meaning: cooking fuel deprivation was most likely the most common of all the indicators, while asset deprivation was most likely the least common of all indicators. One advantage of modeling the other aspects of the data as we have is that we better estimate the uncertainty about easily believable parameters such as the indicator prevalences.

In general, it is wise to always keep in mind what insights the model readily affords and what insights require justification based on elements beyond the scope of the model.

Regression approaches for more complex models

So far, we have elaborated on a unidimensional CFA model showing its connection to regression modeling. And we have borrowed knowledge from both regression and SEM fields to answer questions about the living standards data. Additionally, we have presented heuristic approaches for setting priors on parameters that rely on familiarity with the response scale for the data. We expect that the approaches presented above will be informative for addressing a wide range of SEM problems.

We now devote the next few sections to briefly presenting the regression formulation for a variety of common SEM models, see Table 2 for a summary.

Table 2
Regression equations for commonplace SEM models

Model name	Equation for response y given by respondent r to indicator i , using long-form data
Congeneric model	$y_{ir} \sim \mathcal{N}(\alpha + \beta_i + \lambda_i \theta_p, \sigma_i), \quad \theta_p \sim \mathcal{N}(0, 1)$
MIMIC model with: 2 predictors x_{p1}, x_{p2}	$y_{ir} \sim \mathcal{N}(\alpha + \beta_i + \lambda_i \theta_p, \sigma_i),$ $\theta_p \sim \mathcal{N}(\delta_1 x_{p1} + \delta_2 x_{p2}, 1)$
MIMIC (Alternative formulation)	$y_{ir} \sim \mathcal{N}(\alpha + \beta_i + \lambda_i \theta'_p + \delta'_1 x_{p1} + \delta'_2 x_{p2}, \sigma_i),$ $\theta'_p \sim \mathcal{N}(0, 1)$
Multidimensional CFA with 3 factors	$y_{ir} \sim \mathcal{N}(\alpha + \beta_i + \lambda_{if} \theta_{pf}, \sigma_i),$ $\theta_{pf} \sim \mathcal{N}_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix} \right)$
SEM (latent regressors only): Regress factor 3 on factors 1 and 2 with coefficients γ_1 & γ_2 respectively	Same model as multidimensional CFA, solve: $\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{31} \\ \rho_{32} \end{bmatrix}$
Full SEM: Regress factor 3 on factors 1 and 2 and predictors: x_{p1}, x_{p2}	Same model as multidimensional CFA, but modify covariance matrix of random effects to include observed predictors: $\left\{ \theta_{p1}, \theta_{p2}, \theta_{p3}, x_{p1}, x_{p2} \right\} \sim \mathcal{N}_5(\boldsymbol{\mu}, \mathbf{D}\mathbf{R}\mathbf{D})$ where $\boldsymbol{\mu} = [0, 0, 0, m_1, m_2]$ $\mathbf{D} = [1, 1, 1, s_1, s_2] \times \mathbf{I}_5;$ $m_1(s_1)$ and $m_2(s_2)$ are means (and SDs) of $x_{p1},$ x_{p2} such that $\mathbf{D}$ is a diagonal matrix of SDs, and $\mathbf{R}$ is the correlation matrix with elements: $\rho_{..}$ $\text{Solve: } \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \rho_{14} & \rho_{15} \\ \rho_{21} & 1 & \rho_{24} & \rho_{25} \\ \rho_{41} & \rho_{42} & 1 & \rho_{45} \\ \rho_{51} & \rho_{52} & \rho_{54} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{31} \\ \rho_{32} \\ \rho_{34} \\ \rho_{35} \end{bmatrix}$

Note. MIMIC is the multiple indicators multiple cause model. $\mathbf{I}_5$ is the identity matrix of dimension 5. Solving for γ is easy in commonplace Bayesian programs such as BUGS and JAGS (`inverse(A)%*%b`) and Stan (`mdivide_left_spd(A,b)`), where $\mathbf{A}$ is the covariance matrix to be inverted and $\mathbf{b}$ is the vector of covariances between the predictors and outcome.

Multiple indicator multiple cause (MIMIC) model

The multiple indicator multiple cause (MIMIC) model (Joreskog & Goldberger, 1975) assumes that a factor has one or more *causes* in addition to indicators. Table 2

shows a sample MIMIC model with two observed predictors which vary at the level of the respondent. As shown in Table 2, the mean of the latent variable is assumed to vary as a function of the predictors. A sample predictor in our living standards data would be region of the world for each country.

We also present an alternative formulation for the model that we expect would result in better estimation efficiency. However, this formulation does not suggest the MIMIC framing of the observed predictors as *causes* of the latent variable. The first equation follows the two-level equation approach while the second formulation follows the *mixed* equations in hierarchical models (Raudenbush & Bryk, 2002).

Multidimensional CFA

From the viewpoint of the regression specification, this model is an extension of the congeneric model with two notable differences. First, there are as many random effects as there are hypothesized factors, and the random effects are assumed multivariate normal. Also, the relation between random effects and the responses to items (loadings) vary by both the item and the factor. If one constrains the loadings to be the same within each factor, these models can be estimated using standard multilevel regression software.

Structural models that include only latent regressors

This model is mathematically equivalent (MacCallum, Wegener, Uchino, & Fabrigar, 1993) to the multidimensional CFA. Conceptually, the model differs from the multidimensional CFA in that factors 1 and 2 predict factor 3. Given the equivalence of this model to the multidimensional CFA, one can retain the regression specification from the multidimensional CFA with the same choice of priors. However, there is still need to obtain the coefficients. Using the example in Table 2, these would be the coefficients of factors 1 (γ_1) and 2 (γ_2) predicting factor 3. Table 1 contains the relevant equation for solving for γ_1 and γ_2 .

Given our use of Bayesian estimation, solving this equation for every posterior iteration allows us to obtain posterior samples for the coefficients. Hence, one can still

quantify the uncertainty in the parameters, γ_1 and γ_2 . Commonplace Bayesian engines like BUGS, JAGS and Stan will allow the user to perform this computation within the Bayesian program as shown in the notes in Table 2.

This approach is a two-step process: Bayesian modeling followed by a calculation or transformation of the random effect correlation matrix. A pedagogical benefit of this two-step process is that it centers the fact that a structural model in the SEM framework is no different from simply correlating the relevant variables.

Structural models that include both latent and observed variables

The final model in Table 2 extends the immediate previous model and the MIMIC model by permitting both observed and latent variables as predictors of a factor.

The regression presentation for this model retains the same strategy from the previous model with slight modifications. The observed predictors do not appear in the equation for the item-responses; they only appear in the computation of the covariance matrix that was reserved for random effects.

The covariance matrix can be decomposed in the form $\mathbf{D}\mathbf{R}\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix of standard deviations. The standard deviations of the latent variables are set to 1 and those for the observed predictors are estimated from the data. The correlation matrix of the latent and observed variables that comprise the structural model, $\mathbf{R}$, can then be used to compute the regression coefficients as described in Table 2.

Discussion

We have laid out regression formulations for commonplace SEM models. We hope that laying out the models in this manner provides some clarity about these models such that our language about these models are grounded in meaningful rather than rote learning. Additionally, we hope the regression specifications allow researchers and practitioners think more creatively about these models. For example, as we demonstrated using OLREs and from the regressions in Table 2, it is clear that one can include predictors that vary at the level of the both the indicator and the respondent.

An example of such a predictor would be an individual's response time to an item on a test or survey, allowing for the possibility of adjusting model estimates for responses made in error. This extension is not possible if one uses standard SEM software as item predictors cannot vary within a respondent, but a regression specification places no restrictions on the nature of predictors.

Another immediately obvious extension the regression approach allows for is the extension of SEM models to easily accommodate multiple nested or crossed random effects (which are common in education). The GLLAMM framework (Rabe-Hesketh et al., 2004) leverages this freedom. For example, researchers may administer a test to a large number of students; the students are within classrooms within schools within districts, and the students may come from different neighbourhoods. Depending on available data, the regression specification allows for specifying a simple random intercept or a more elaborate model at each grouping level. At a minimum, specifying a random intercept (or a parallel indicator CFA) at incidental levels would be helpful for estimating the uncertainty about parameters of interest. If one is willing to assume that all the item loadings are identical within each grouping level (parallel model), then one can estimate a multilevel CFA accounting for multiple incidental levels using commonplace multilevel software (see Li, Duncan, Harmer, Acock, & Stoolmiller, 1998, for an illustration with HLM software).

We have approached specifying SEMs as regression using Bayesian estimation. As we demonstrated, Bayesian estimation allows for the easy estimation of congeneric models where loadings are allowed to vary by item. Additionally, general purposes Bayesian computation engines provide a wide variety of statistical distributions (and users may specify custom distributions in Stan) providing great flexibility in how we model variables. As seen in our example with the living standards data, the choice of distribution can have a consequential impact on model insights.

Additionally, the Bayesian approach allows for answering questions probabilistically. For example, we were able to rank the prevalence of the different living standards indicators probabilistically using the posterior distributions of the

indicator means. Analyzing data in this manner broadens the array of research questions we can ask and how we answer them. However, there can be some challenges when adopting a Bayesian approach, such as prior specification and the computational burden of writing model code. It is for these reasons that we share the Stan and R code associated with the models in this paper. We expect that given some familiarity with Stan, readers will be able to modify code for their problems. And we hope that our treatment of prior specification provides sufficient guidance for researchers and teachers when modeling. This guidance will be very relevant for researchers working with the small sample sizes (Smid, McNeish, Miočević, & van de Schoot, 2020).

Our modeling approach here has implications for the discussion of whether to treat item parameter estimation as random or fixed (De Boeck, 2008). In the item-response literature, random item models are suggested as useful when the “items are sampled from a larger population of items” (Raaijmakers, Schrijnemakers, & Gremmen, 1999, p. 416). An examination of the regression models presented in this paper shows that the indicator means were estimated as random. However, the indicators under consideration are uniquely interesting indicators of living standards, perhaps suggesting a fixed effects approach. We would argue that a random items approach should be standard regardless of whether the items are sampled from a larger population or uniquely interesting. Our justification is the same justification for estimating the country averages as random: we treat the indicators means as random effects because of the estimation efficiency benefits. Moreover, this justification avoids the confusion associated with common advice on when to model parameters as fixed or random.⁷

We have demonstrated SEM as regression using bounded continuous data, covered the commonplace Gaussian CFA and not-so-common beta CFA. Our treatment here has not covered SEMs for discrete data, e.g. binary and ordinal items. However, the models in Table 2 can be modified slightly to accommodate discrete types. The biggest changes when adapting the specifications in Table 2 to binary, ordinal or count items would

⁷ (Gelman & Hill, 2007, p. 245) document diverging recommendations from methodologists on whether to estimate parameters as fixed or random. It is often cleaner to adopt the view that random effects are shrinkage devices that improve estimation efficiency, and the normal prior on the parameters is an analytically convenient choice.

relate to the specification of error variances. Moreover, by treating the data as beta-distributed, we have provided some insights into how one might model data on a scale that is different from the response scale. For the beta-distributed data, we assumed the mean of the data to be linear on logit-scale, and a similar assumption is often made about binary and ordinal items.⁸ Thus some of the advice here on prior specification is readily applicable to Bayesian regression treatment of binary and ordinal SEMs.

Open Practices Statement

All of the data and scripts for the analyses and graphs reported here are available at: https://github.com/jamesuanhoro/sem_mlm_brm.

⁸ We say similar because SEM approaches in software for binary/ordinal data often rely on estimation of tetrachoric/polychoric correlations, thus assuming the relations are linear on the probit- rather than logit- scale (Ronning & Kukuk, 1996).

References

- Alkire, S., Kanagaratnam, U., & Suppa, N. (2018). *The Global Multidimensional Poverty Index (MPI): 2018 Revision* (MPI Methodological Notes No. 46). Oxford, England: OPHI. Retrieved from <https://ora.ox.ac.uk/objects/uuid:eaac77b5-4398-4588-a456-4996f6e147ba>
- Birnbaum, A. (2008). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Information Age Publishing. (Original work published in 1968)
- Bollen, K. A. (1989). *Structural equations with latent variables*. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi: 10.1002/9781118619179
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. doi: 10.32614/RJ-2017-066
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). doi: 10.18637/jss.v076.i01
- Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78(4), 685–709. doi: 10.1007/s11336-013-9328-2
- Curran, P. J. (2003, oct). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38(4), 529–569. doi: 10.1207/s15327906mbr3804_5
- Davis-Stober, C. P., Dana, J., & Rouder, J. N. (2018). Estimation accuracy in the psychological sciences. *PLOS ONE*, 13(11), e0207239. doi: 10.1371/journal.pone.0207239

- De Boeck, P. (2008, dec). Random item IRT models. *Psychometrika*, 73(4), 533–559.
doi: 10.1007/s11336-008-9092-x
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach* (P. De Boeck & M. Wilson, Eds.). New York, NY: Springer New York. doi: 10.1007/978-1-4757-3990-9
- Falk, C. F. (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling*, 25(2), 244–266. doi: 10.1080/10705511.2017.1367254
- Ferrari, S., & Cribari-Neto, F. (2004, aug). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. doi: 10.1080/0266476042000214501
- Frank, S. A. (2009, aug). The common patterns of nature. *Journal of Evolutionary Biology*, 22(8), 1563–1585. doi: 10.1111/j.1420-9101.2009.01775.x
- Gelman, A., Goegebeur, Y., Tuerlinckx, F., & Van Mechelen, I. (2000). Diagnostic checks for discrete data regression models using posterior predictive simulations. *Applied Statistics*, 49(2), 247–268. doi: 10.1111/1467-9876.00190
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4), 733–760. Retrieved from <http://www.jstor.org/stable/24306036>
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972, sep). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42(3), 237–288. doi: 10.3102/00346543042003237
- Greenland, S., & Mansournia, M. A. (2015). Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Statistics in Medicine*, 34(23), 3133–3143. doi: 10.1002/sim.6537
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation*

and multilevel modeling approaches.

- Harrison, X. A. (2014, oct). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2, e616. Retrieved from <https://peerj.com/articles/616> doi: 10.7717/peerj.616
- Hinde, J. (1982). Compound Poisson Regression Models. In R. Gilchrist (Ed.), *Glim 82: Proceedings of the international conference on generalised linear models* (pp. 109–121). New York, NY: Springer New York.
- Hodges, J. S., & Clayton, M. K. (2011). *Random effects old and new* (Tech. Rep.). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.225.2685>
- Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351), 631–639. doi: 10.2307/2285946
- Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(1), 136–153. doi: 10.1080/10705510701758406
- Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Academic Press.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573–603. doi: 10.1037/a0029146
- Li, F., Duncan, T. E., Harmer, P., Acock, A., & Stoolmiller, M. (1998, jan). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(3), 294–306. doi: 10.1080/10705519809540106
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114(1), 185–199. doi: 10.1037//0033-2909.114.1.185
- Maul, A. (2017, apr). Rethinking traditional methods of survey validation.

- Measurement: Interdisciplinary Research and Perspectives*, 15(2), 51–69.
Retrieved from <https://doi.org/10.1080/15366367.2017.1348108> doi:
10.1080/15366367.2017.1348108
- Mayer, R. E. (2002, nov). Rote versus meaningful learning. *Theory Into Practice*, 41(4), 226–232. Retrieved from https://doi.org/10.1207/s15430421tip4104_4 doi:
10.1207/s15430421tip4104_4
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press.
- O’Connell, A. A., Yeomans-Maldonado, G., & McCoach, D. B. (2015). Residual diagnostics and model assessment in a multilevel framework: Recommendations toward best practice. In J. Haring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 97–135).
- Piironen, J., & Vehtari, A. (2017, may). Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3), 711–735. doi:
10.1007/s11222-016-9649-y
- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “The Language-as-Fixed-Effect Fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41(3), 416–426. doi:
10.1006/jmla.1999.2650
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004, jun). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167–190. doi:
10.1007/BF02295939
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.
- Ronning, G., & Kukuk, M. (1996, jun). Efficient estimation of ordered probit models. *Journal of the American Statistical Association*, 91(435), 1120–1129. doi:
10.2307/2291731
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of*

- Statistical Software*, 48(2), 1–36. doi: 10.18637/jss.v048.i02
- Rovine, M. J., & Molenaar, P. C. M. (2000, jan). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, 35(1), 51–88. doi: 10.1207/S15327906MBR3501_3
- Ruppert, D. (1987, may). Robust statistics: The approach based on influence functions. *Technometrics*, 29(2), 240–241. doi: 10.1080/00401706.1987.10488218
- Savalei, V. (2014, jan). Understanding robust corrections in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(1), 149–160. doi: 10.1080/10705511.2013.824793
- Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling*, 27(1), 131–161. doi: 10.1080/10705511.2019.1577140
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71. doi: 10.1037/1082-989X.11.1.54
- Song, X. Y., & Lee, S. Y. (2012). *Basic and advanced Bayesian structural equation modeling*. Wiley. doi: 10.1002/9781118358887
- Stan Development Team. (2016). *CmdStan: User's guide*. Retrieved from <https://github.com/stan-dev/cmdstan/releases/download/v2.23.0/cmdstan-guide-2.23.pdf>
- Takane, Y., & de Leeuw, J. (1987, sep). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52(3), 393–408. doi: 10.1007/BF02294363
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Burkner, P.-C. (2020). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 1–28. Retrieved from <https://projecteuclid.org:443/euclid.ba/1593828229> doi: 10.1214/20-BA1221

West, M. (1984, aug). Outlier Models and Prior Distributions in Bayesian Linear Regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3), 431–439. Retrieved from <http://www.jstor.org/stable/2345685>

Appendix

Figure 14. Rank plots for parameters from Bayesian congeneric CFA assuming normality.

Figure 15. Rank plots for parameters from Bayesian congeneric CFA (with OLRE) assuming beta distribution.

Figure 16. Posterior predictive checking by overlaying data simulated (100 simulations) from normal distribution model against the density of original data (solid black line).

Figure 17. Posterior predictive checking by overlaying data simulated (100 simulations) from beta congeneric CFA in equation 9 (no OLRE) against the density of original data (solid black line).