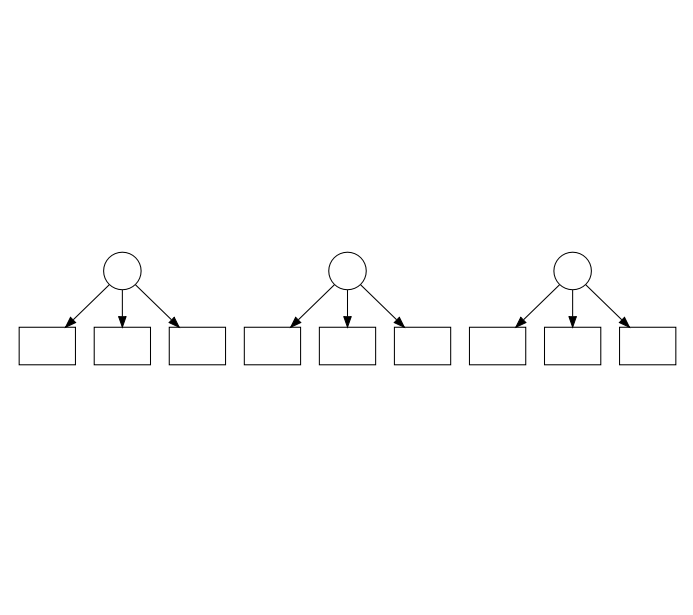
When specifying the measurement component of a structural equation model (SEM) careful thought has to be given to the specific pattern of factor loadings that will be estimated. A number of concerns will motivate this decision; the loading pattern should reflect substantive theory, there should be sufficient flexibility to allow for unexpected findings to arise from data, the estimated solution should adequately fit the observed data, and the practical concerns of model identifiability must be met. Unfortunately these concerns are often in conflict with one another, for example a more flexible specification may complicate the relation between model and theory (Stromeyer, et al., 2015). As a result it is necessary that some compromise is reached between such concerns. How this compromise can best be reached is an open question, with an increasing recognition that traditional approaches may not provide an optimal balancing of these concerns.

The independent clusters (IC) factor loading pattern, where only a single factor loading is estimated for each observed variable, represents the most common compromise between the above concerns in SEM. Figure 1. provides an example diagram and loading matrix for an IC structure with three latent and nine observed variables. The majority of loading parameters are fixed at zero, with only a relatively small proportion freely estimated.

FIGURE 1 GOES HERE



By only estimating a single loading per observed variable the IC structure ensures model identifiability under reasonable conditions, allows for straightforward mapping of theory on to the model, and provides easily interpretable results. However, it achieves this at the cost of model flexibility, better fitting loading patterns may exist but the IC structure is unable to speak to these due to its enforced rigidity. Due to this, the IC structure is increasingly being recognised as an unsatisfactory solution to the problem of specifying effective factor loading patterns (Asparouhov & Muthén, 2009; Muthén & Asparouhov, 2012).

Muthén and Asparouhov (2012) propose Bayesian Structural Equation Modelling (BSEM) as a method for simultaneously estimating all possible factor loadings in SEM. Their method overcomes the above limitations of IC structure by allowing for any loading estimate to differ from zero if the observed data suggests it, greatly increasing flexibility in SEM loading specification. This is achieved by setting informative Bayesian priors, specifically small variance Gaussian distributions with mean zero, on all cross-loading parameters, rather than fixing them to zero as in IC structure. This allows for estimated cross loadings to differ from zero whilst maintaining model identifiability.

However BSEM, as proposed by Muthén and Asparouhov, achieves this at the expense of the clear relationship between model and theory that IC models promote. As Stromeyer, et al. (2015) discuss in their critique of the method, the use of Gaussian priors in this manner can result in a complex pattern of cross-loadings, with sampling error alone resulting in a large number of apparently non-zero cross loading parameters. This not only complicates the interpretation of the latent variables but also makes one goal of BSEM, to systematically identify reliably non-zero cross-loadings, more difficult as some decision criteria must then be chosen to separate ‘significant’ from ‘non-significant’ cross loadings.

BSEM would therefore seem to free SEM from a restrictive approach to cross-loading estimation only at the expense of losing the direct and easy to interpret relationship between theory and model that makes SEM such a productive tool. However, the literature on Bayesian regularisation, concerned with modelling sparse parameter sets where most population values are equal to zero, may provide a means to achieve systematic cross-loading estimation whilst maintaining a sparse, easy to interpret cross-loading pattern similar to IC structure.

This paper investigates the potential for using informative priors common in the Bayesian regularisation literature, instead of Gaussian priors, to estimate SEM cross-loadings. It was hoped that the use of sparsity promoting priors would provide a better compromise between model flexibility and interpretability than either BSEM with Gaussian priors or IC models currently provide.

A brief overview of the principles of Bayesian regularisation is first provided, and candidate prior distributions are introduced. Simulations comparing performance of small variance Gaussian priors against two priors from the regularisation literature, the horseshoe and horseshoe plus, are then reported and discussed. Finally, the impact of prior distribution selection on cross loading estimation is explored using the Holzinger and Swineford (1939) ability testing dataset.

**Bayesian Regularisation and Sparse Estimation**

Regularisation describes a range of approaches to model estimation which broadly seek to simplify the estimated parameter solution, subject to some penalty on model complexity (Hastie, Tibshirani, & Friedman, 2001). This almost always entails certain parameter estimates being ‘shrunk’ towards zero, relative to a non-regularised solution, in order to better meet the conditions of the penalty. For this reason regularisation is also often referred to as shrinkage modelling.

Regularisation models have gained popularity in disciplines where complex models are at risk of overfitting sample data, or where a large number of candidate predictor variables make traditional means of variable selection (e.g. null hypothesis significance testing) untenable. Certain regularisation techniques can also result in improved performance in poorly posed modelling situations, such as in the presence of high multicollinearity (Vinod, 1978). They can also enable the estimation of models which would have been unidentified without the application of regularisation, such as regression models where candidate predictor variables outnumber data observations (Zou & Hastie, 2005).

Within a frequentist framework regularisation is achieved by adding some penalty on model complexity to the relevant loss/likelihood function. For example, a linear ridge regression model is estimated via ordinary least squares (OLS) plus an additional penalty that is proportional to the sum of squared regression coefficients (Hoerl & Kennard, 1970), as below.

The parameter λ is a tuning parameter which controls the strength of regularisation to be applied. When λ equals 0 the model reduces to basic OLS regression whilst as it increases all relevant coefficients will be asymptotically shrunk to zero. The value of λ in frequentist regularisation is typically selected using cross-validation (Tibshirani, 1996). Different specifications of the penalty term will result in different shrinkage behaviour. For example, LASSO regularisation uses a linear, rather than quadratic, penalty (Tibshirani, 1996) whilst Elastic Net regularisation uses a mixture of linear and quadratic penalties (Zou & Hastie, 2005).

Similar results are achieved in a Bayesian context by setting tight prior distributions, centred on zero, on the relevant model parameters (typically regression coefficients). By specifying prior distributions with a high probability density near zero, equivalent shrinkage of posterior estimates towards zero can be achieved, bringing with it the benefits associated with frequentist regularisation described above (Park & Casella, 2008). So long as they have a high probability density near zero, a variety of different distributions can be used to achieve different regularisation properties, similar to the use of different complexity penalties in frequentist regularisation.

Specific prior distributions have been shown to produce regularised estimates closely related to specific frequentist regularisation methods. For example the frequentist LASSO is broadly equivalent to a Bayesian model in which Laplace distributions are set as priors on the relevant parameters (Park & Casella, 2008). Importantly for the present discussion, Bayesian ridge regression is achieved by setting small variance Gaussian priors on the regression coefficients (Pasanen, Holmström, & Sillanpää, 2015). The BSEM approach to cross loadings advocated by Muthén and Asparouhov (2012) can therefore be interpreted as a use of ridge regression, applied only to cross loading parameters, to overcome the problem of identifiability when all SEM loadings are freely estimated.

This perspective casts the critique of Stromeyer, et al. (2015) in a productive light. Their major concerns with BSEM, using small variance Gaussian priors, relate to the fact that models estimated in this way will commonly display a relatively complex pattern of cross loading estimates. They raise the issue that having such a large number of, theoretically secondary, loadings estimated at non-zero values harms the interpretability of the estimated model. They also raise the concern that the observed tendency of estimated cross loading posteriors to follow the data away from zero, could complicate model falsifiability with even poorly stated models achieving reasonable fit through an overfitting of cross-loading parameters.

These issues arise from the behaviour of ridge regression as a regularisation technique, specifically that it does not promote sparse parameter solutions. This is best seen by comparison with LASSO regression, which does promote sparsity. **Figure 2** shows, on a simulated dataset, how LASSO and ridge regression coefficient estimates change as the regularisation penalty parameter is increased. Even for very large values of λ ridge regression produces non-zero estimates for almost all variables. The LASSO by comparison quickly shrinks most estimates to zero, retaining only a sparse set of non-zero estimates for most values of λ.

FIGURE 2 GOES HERE

A potential development of BSEM, in response to the concerns of Stromeyer, et al. (2015), is therefore to explore the use of regularisation priors that promote a sparse parameter solution, rather than the noisy solution provided by the use of Gaussian priors (comparable to ridge regression). Such a sparse solution could not only aid interpretability of the estimated model, but could also help avoid the overfitting of cross-loadings near zero that the use of Gaussian priors may invite. The literature on Bayesian regularisation provides several such candidate priors.

**Sparsity Promoting Prior Distributions**

A number of prior distributions can be used to promote sparsity in Bayesian estimation. As shown above, the LASSO is one such technique. The Bayesian LASSO is achieved by setting a Laplace prior distribution on the parameters to undergo regularisation (Park & Casella, 2008). The Laplace distribution can be parameterised in a number of equivalent forms, but is sometimes referred to as the double exponential distribution (Andrews & Mallows, 1974) as it can be expressed as two exponential distributions placed back to back, centred over zero.

In addition to the Laplace, the Horseshoe prior (Carvalho, Polson, & Scott, 2009) is a recent development, which promotes sparsity more powerfully than the LASSO but is also better placed to handle substantive non-zero values due to its fatter tails. Following the notation of Carvalho, Polson and Scott (2009), the Horseshoe distribution can be described as a Gaussian distribution whose variance is the product of two other random variables. If represents a set of values distributed according to a horseshoe distribution, then their distribution can be written as follows:

Here represents the Cauchy plus distribution, a Cauchy distribution limited to non-negative values (Gelman, 2006). By using both ‘global’ and ‘local’ parameters to construct the horseshoe distribution it allows for flexibility in the estimation of sparse parameter sets that is lacking from the Laplace distribution. Comparisons tend to show use of horseshoe priors outperforming the Bayesian LASSO in such cases (Peltola, Havulinna, Salomaa, & Vehtari, 2014).

A further development on the horseshoe prior is the horseshoe+ prior, which was explicitly designed to provide more optimal estimation of sparse parameter sets than competitors, including the horseshoe and Laplace distributions (Bhadra, Datta, Polson, & Willard, 2015). The horseshoe+ is defined similarly to the horseshoe distribution, but with an additional ‘local’ half Cauchy mixture component in the variance term.

The Laplace, horseshoe, and horseshoe+ distributions were all considered as alternatives to the Gaussian priors of Muthén and Asparouhov (2012) in the below simulations. **Figure 3** shows how these distributions differ from one another, both at the centre of the distribution and in the tails[[1]](#footnote-1). The key properties they share are large probability volume around the centre of the distribution, and fat tails by comparison with the Gaussian distribution. The high density near zero promotes sparsity by strongly shrinking small estimates towards zero. The fat tails help reduce shrinkage of valuesthat are further from the central value.

FIGURE 3 GOES HERE

**Simulation Study of Sparsity Promoting Priors**

In order to compare different cross-loading priors, data were simulated from two population models. One where there are a small number of non-zero cross loadings, and one where all population cross-loadings were zero (that is, the IC structure is accurate). As researchers will likely not know whether there are any non-zero cross-loadings in advance of modelling, this provides a fuller evaluation of the performance of BSEM in realistic situations.

Both datasets were simulated using the Simsem R package (Pornprasertmanit, Miller, & Schoemann, 2015) from a population model with 15 observed variables with main loadings on to 3 latent variables, with values between 0.4 and 0.8. Latent variables were uncorrelated and observed variables were all conditionally independent given the latent variables. All latent and observed variables had variances of 1.

The only difference between the IC population model and the sparse cross-loadings (SCL) model were that in the former all cross-loadings were equal to 0, whilst in the SCL model 3 cross-loadings were given non-zero population values. The two loading matrices are shown in figure 4. 100 datasets, each with n = 200, were simulated for both the IC and SCL models.

FIGURE 4 GOES HERE

**IC Factor Matrix**  **SCL Factor Matrix**

The simulation study aimed to evaluate model behaviour on four points:

1. **The ability to estimate population parameters with minimal bias.** It was expected for cross-loadings with a population value of zero that all priors would perform equivalently in returning un-biased estimates. However it was expected that for non-zero values the Gaussian prior would show larger bias towards zero than the alternative distributions due to its thinner tails.
2. **The variance of estimates from sample to sample.** As it does not promote sparsity, the Gaussian prior was expected to show considerably greater variation in its estimates from sample to sample, particularly in its estimates of cross-loadings with population values of zero.
3. **Effect of prior choice on model fit.** It was expected that the Gaussian priors would provide poorer model fit than the alternatives, due to larger bias of estimates for non-zero parameters, and due to overfitting of near zero values.
4. **Model Complexity**. It was expected that the Gaussian prior would result in a more complex model than the alternatives (as measured by the effective number of parameters), due to its non-sparse solutions.

A direct comparison of model interpretability is provided in a later section using a single non-simulated dataset.

**Modelling Information**

All BSEM models were estimated using the Stan probabilistic programming language (Carpenter, et al., in press) from within R using the RStan package. Stan was chosen as its use of Hamiltonian Monte Carlo (implemented using the No-U-turn sampler) tends to provide better sampling behaviour than alternative languages which commonly make use Metropolis-Hastings or Gibbs sampling (Hoffman & Gelman, 2014).

Gaussian priors with a mean of 0 and variance of 3 were set on the primary factor loadings (these were presumed to be known in advance); these parameters were also restricted to positive values to achieve model identification. Whilst their population values equal zero, latent variable correlations were estimated to allow for sampling variation. Observed variables were modelled as conditionally independent without correlated residuals. All fitted models were identical except for the prior distributions set on the cross-loading parameters. The Stan code for each model can be found at (GITHUB GOES HERE).

All models were explored individually to ensure that the estimation process converged reliably in each case. This meant that the number of MCMC iterations ran, the thinning parameters used, and the Metropolis acceptance step parameters, differ somewhat between models, depending on how well the MCMC process sampled from the posterior in each case. In all cases between 1000 and 2000 MCMC iterations were ran with a burn in of 50% of samples. Thinning was between three and five cases. Reliable convergence was identified in all cases before simulations were run. Full details are recorded in the simulation code which can be found at (GITHUB GOES HERE).

Approximate leave-one-out cross-validation (LOO) was used a measure of model fit for comparison (Vehtari, Gelman, & Gabry, 2016). LOO is a recently proposed measure of Bayesian model fit which asymptotically estimates full leave-one-out cross validation but without the need to carry out the computationally expensive process of cross validation. This is achieved through log-likelihood evaluations using samples from the fitted posterior distribution (full details are reported in Vehtari, Gelman and Gabry, 2016). The strength of LOO is it’s ability to produce out of sample fit estimates (helping to identify and avoid overfitting)whilst avoiding identified issues with alternative such as the deviance information criterion (DIC; Plummer, 2008). LOO was calculated for all models after estimation using the loo package for R. This package was also used to estimate the effective number of parameters for each model in order to compare model complexity.

**Laplace prior estimation**

During both the initial tests of model convergence, and later when the full simulation study was ran, issues were identified with the convergence of models using the Laplace prior distribution on it’s cross loading parameters. Even when using a large set of MCMC samples, and taking steps such as increasing the thinning parameters, convergence of these models was found to be poor, with non-convergence more common than not.

As reliable convergence could not be achieved the Laplace prior is not considered further below. It was considered that this failure was worth reporting as the LASSO method which the Laplace prior achieves is the most popular method of sparsity promoting regularisation. As such it’s exclusion from consideration entirely would have been peculiar.

**Sparse Cross Loading Comparison**

Simulations were first ran on the SCL datasets to compare the performance of different prior values in the presence of a small number of non-zero cross loadings. Figures 5 - 7 display the median and 95% coverage interval of all cross-loading parameter estimates over the 100 simulated datasets (the parameter estimate here is the posterior median) for each prior distribution.

FIGURES 5 – 7 GO HERE

The horseshoe and horseshoe+ priors both reliably produced near exact estimates of zero where the population value was zero. Across almost all simulated models they resulted in an appropriately sparse solution to help mitigate potential difficulties of model interpretation. The Gaussian prior, whilst providing a median estimate across all models very near to zero, displayed a high level of variation from sample to sample, highlighting it’s susceptibility to over fit estimates based on sampling variation.

The horseshoe and horseshoe+ priors also produced a less biased estimate of the cross-loadings with population values of .6 and .4 than the Gaussian. Due to their relatively fat tails the horseshoe and horseshoe+ estimates were shrunk more minimally than with the thin tailed Gaussian. However, it’s clear that the non-Gaussian priors struggle with the cross-loading with population value of .2, with the median estimate across samples lying very close to zero. This results from the relatively small sample sizes (n=200) and the strong shrinkage behaviour of these priors near zero. Due to its weaker shrinkage behaviour the Gaussian prior tended to provide a less biased estimate of the true population value. This weakness of the non-Gaussian priors could be overcome with larger sample sizes.

In order to study the effects of prior choice on model fit a multilevel model was used, at this stage models were also fit to the 100 simulated datasets which had cross-loading priors fixed at 0, producing strict independent clusters structure. This was to act as a baseline for model fit. All 400 estimated LOO values were regressed firstly on to an indicator variable for which dataset each model was fit to, thus accounting for variance between simulated datasets (these parameters were modelled as random effects). Against this baseline model was compared a model that, in addition to using a dataset indicator variable, added a fixed indicator of which prior distribution was set on the cross-loading parameters. A likelihood ratio test of the two models showed that the choice of priors had a significant impact on model fit, χ2 (3) = 268.19, p < .001.

Figure 8 shows the coefficients for each prior distribution from the above model, with the independent clusters model acting as a baseline. Consistent with the results of Muthén & Asparouhov (2012) the use of Gaussian priors on cross-loadings provides notably better fit than independent clusters structure when there are cross-loadings not equal to zero. However, both the horseshoe and horseshoe+ priors can be seen to provide just as notable an increase in model fit over the Gaussian prior models as they do over the independent clusters model. This is consistent with the hypothesis that larger bias for non-zero values, and a tendency to overfit based on sampling variation, would harm the LOO of models using Gaussian priors relative to the alternatives.

FIGURE 8 GOES HERE

The final aspect on which models were compared was their complexity, as measured by their LOO effective number of parameters (see Vehtari, Gelman, & Gabry, 2016, for full details on the calculation of this value). As with LOO fit a multilevel model was employed which regressed the effective number of paramaters on to a dataset indicator to control for variation between simulated datasets. This model was then compared against one in which indicators for cross-loading priors were included (no intercept term was included in either model to allow direct comparison between prior distributions) , a likelihood ratio test showed that inclusion of prior distribution indicators significantly improved model fit, χ2 (4) = 863.1, p < .001.

Figure 9 shows the coefficients for each prior distribution when modelling the effective number of parameters (with no intercept term). The independent clusters model is clearly the simplest model with the free estimation of all cross-loadings significantly increasing the number of effective parameters. As expected the use of Gaussian priors results in a more complex fitted model than either the horseshoe or horseshoe+ models. This is a result of weaker shrinkage towards zero and, to some extent, overfitting of estimates that had population values of zero. The effective number of parameters shown here are linked directly to the LOO fit values shown in figure 8, so the difference between the Gaussian and alternative priors not only highlights a difference in fitted model complexity, but also contributes to the difference in LOO fit above.

FIGURE 9 GOES HERE

**Independent Clusters Loading Comparison**

The above process was repeated on the datasets simulated under the assumption that all population cross-loading values were equal to zero. Figures 10 – 12 again display the median and 95% coverage interval of cross-loading estimates over the simulated datasets.

FIGURES 10 – 12 GO HERE

Both methods provide reliable, un-biased median estimates close to zero over all 100 datasets. However it’s clear immediately that this estimate is only a long run property when using Gaussian priors, with individual estimates varying widely between -.2 and .2, following random variation from simulated dataset to dataset. By comparison the strong shrinkage of the horseshoe and horseshoe+ priors result in extremely consistent estimates, returning values indistinguishable from zero in almost all cases.

The analysis of model fit was carried out identically to the above analysis using the SCL datasets. A multilevel model with random dataset indicator variables was estimated, along with a further model which added indicators for cross-loading priors used (including models in which cross-loadings were fixed at zero). A likelihood ratio test of these two models showed that choice of prior distribution had a significant effect on LOO model fit, χ2 (3) = 517.55, p < .001. The fitted coefficients for prior distribution (with priors fixed at zero acting as a baseline) are plotted in figure 13.

FIGURE 13 GOES HERE

Compared to a model in which all cross-loadings are a priori fixed at zero, the use of Gaussian priors results in notably worse model fit in cases where all cross-loadings have population values equal to zero. Interestingly the use of the horseshoe and horseshoe+ priors does not result in a notably worse out of sample model fit than restricting these values at zero. Whilst it might be expected that running a larger number of simulations may allow for a statistically significant increase in model fit to be detected, this finding does suggest that any such effect will be minor next to the worsening of model fit associated with the use of Gaussian priors.

The effect of prior distribution choice, with an independent clusters population structure, on the effective number of model parameters was again approached in the same manner as the above analysis of this question on the SCL datasets. Again, a likelihood ratio test comparing the base model against one which included indicators for cross-loading prior distribution showed a significant improvement in explanatory power with the variables inclusion, χ2 (4) = 1109.7, p < .001. Figure 14 plots the estimated coefficients for the four approaches to cross-loading estimation (again no intercept term was modelled to allow for direct interpretation of effective parameter numbers).

FIGURE 14 GOES HERE

Fixing cross-loadings at zero produced a simpler model than any of the freely estimated alternatives. However, both the horseshoe and horseshoe+ models were had a significantly smaller number of parameters than the Gaussian models. The non-Gaussian priors estimate a model a much simpler model than the Gaussian approach, which assigns greater probability density at values further from zero.

**Discussion of simulation study results**

The effects of cross-loading prior distributions on model fit shown above are considered the most important point of comparison. As expected, in the presence of non-zero cross-loadings there was a significant improvement in fit when these parameters were freely estimated rather than fixed at zero. However, in these cases the use of Gaussian priors significantly underperformed both of the alternative regularisation priors. A combination of greater bias towards zero for large estimates, and overfitting of small estimates near zero, drove this difference in out of sample fit.

Not only did the horseshoe and horseshoe+ priors outperform the Gaussian when there were a small number of non-zero cross-loadings, but they also outperformed it when all cross-loadings had population values of zero. Most notably, when the independent clusters structure was true in the population, no significant difference in model fit was identified between fixing cross-loadings at zero and freely estimating them using the horseshoe or horseshoe+. The use of these regularisation priors at least matches the model fit of independent clusters models when the assumption holds true, and significantly outperforms when it is false.

Both the Gaussian and alternative distributions showed that they were able to fairly reliably identify moderate to large cross-loadings (.4 and .6 respectively). Due to the regularisation being carried out in all cases, through the use of priors with high density around zero, these estimates all showed bias towards zero. However, due to their fatter tails the horseshoe and horseshoe+ priors resulted in smaller overall bias than the use of Gaussians. The one area where the Gaussian priors were seen to outperform the alternatives was in the identification of small (.2) cross loading values. With the sample size used in each simulated dataset (n=200) the large majority of estimates produced using the non-Gaussian priors were shrunk to zero, whilst the use of Gaussian priors did not show this same strong shrinkage.

All three studied priors provided unbiased estimates of near zero cross-loadings over the 100 simulations. However, the Gaussian showed a high degree of variation from sample to sample whereas the horseshoe and horseshoe+ priors reliably returned estimates of exactly zero across almost all samples. When the true parameter structure was simple, with no major cross-loadings, the Gaussian prior resulted in noisy estimates driven by sampling error. By comparison the alternatives reliably identified the correct, simple structure from sample to sample.

These results indicate that the horseshoe and horseshoe+ outperform the Gaussian distribution as a prior on cross-loading parameters, in both individual samples and aggregated over many simulations. On the basis of the above, Stromeyer, et al’s (2015) concern that BSEM would lead to overfitting and consistently complex parameter solutions are considered to be valid when small variance Gaussian priors are used. However, the above simulations show that the strong regularisation of the horseshoe and horseshoe+ distributions avoid these pitfalls. Likewise, concerns over falsifiability are mitigated as the use of the non-Gaussian priors retains a simple structure where appropriate, and avoids the overfitting from sample to sample observed using Gaussian priors.

However, the above simulations can’t speak to one aspect of Stromeyer, et al’s critique, that BSEM models, with cross-loadings freely estimated, will often result in difficult to interpret results due to the large number of loading parameters that must then be considered when interpreting the overall model. Based on the above results it was expected that the use of strong Bayesian regularisation priors would also help minimise this concern by returning relatively simple patterns of cross-loading estimates. This is explored below in a comparison of BSEM prior distributions using a real dataset.

**Comparison of cross-loading priors using the Holzinger and Swineford dataset**

In order to compare the interpretability of BSEM models in a single sample situation the widely studied Holzinger and Swineford (1939) mental abilities dataset was modelled. The models fitted had a loading pattern mirroring that reported in Muthén & Asparouhov (2012), with 19 observed variables regressed on to 4 latent variables. Table 1 lays out this loading structure, observed and latent variable names are those used by Muthén & Asparouhov.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Observed Variable | Spatial | Verbal | Speed | Memory |
| Visual perception | X | 0 | 0 | 0 |
| Cubes | X | 0 | 0 | 0 |
| Paper form board | X | 0 | 0 | 0 |
| Flags | X | 0 | 0 | 0 |
| General information | 0 | X | 0 | 0 |
| Paragraph comprehension | 0 | X | 0 | 0 |
| Sentence completion | 0 | X | 0 | 0 |
| Word classification | 0 | X | 0 | 0 |
| Word meaning | 0 | X | 0 | 0 |
| Addition | 0 | 0 | X | 0 |
| Code | 0 | 0 | X | 0 |
| Counting groups of dots | 0 | 0 | X | 0 |
| Straight and curved capitals | 0 | 0 | X | 0 |
| Word recognition | 0 | 0 | 0 | X |
| Number recognition | 0 | 0 | 0 | X |
| Figure recognition | 0 | 0 | 0 | X |
| Object-number | 0 | 0 | 0 | X |
| Number-figure | 0 | 0 | 0 | X |
| Figure-word | 0 | 0 | 0 | X |

Apart from the changes in observed and latent variable numbers, and the associated new loading structure, the BSEM models fit are exactly as described in the above simulation studies, with uncorrelated observed variable errors, main loading priors of N(0, 3) and only cross-loading priors differing between models. In both this study, and the above simulations, the horseshoe and horseshoe+ distributions showed no substantive differences in estimation. For this reason only the models using Gaussian and horseshoe priors are reported in the below comparison. Full Stan code for all models can be found at (GITHUB GOES HERE). In this comparison only the data from the Grant-White school (N=156) was used.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Gaussian priors | | | |  | Horseshoe priors | | | |
| Observed Variable | Spatial | Verbal | Speed | Memory |  | Spatial | Verbal | Speed | Memory |
|  |  | Loadings (posterior median) | | | | | | | |
| Visual perception | **.66\*** | -.01 | .12 | .12 |  |  |  |  |  |
| Cubes | **.64\*** | .10 | -.01 | -.10 |  |  |  |  |  |
| Paper form board | **.50\*** | .13 | .10 | -.14 |  |  |  |  |  |
| Flags | **.60\*** | .10 | -.05 | .11 |  |  |  |  |  |
| General information | .06 | **.83\*** | -.06 | -.07 |  |  |  |  |  |
| Paragraph comprehension | .00 | **.80\*** | -.07 | .11 |  |  |  |  |  |
| Sentence completion | -.05 | **.93\*** | .00 | -.02 |  |  |  |  |  |
| Word classification | .11 | **.69\*** | -.01 | .07 |  |  |  |  |  |
| Word meaning | -.01 | **.94\*** | -.02 | -.09 |  |  |  |  |  |
| Addition | -.26\* | .10 |  | .02 |  |  |  |  |  |
| Code | .04 | -.01 |  | .24\* |  |  |  |  |  |
| Counting groups of dots | .06 | -.01 |  | -.11 |  |  |  |  |  |
| Straight and curved capitals | .29\* | -.18 |  | -.02 |  |  |  |  |  |
| Word recognition | -.16 | .09 | -.13 |  |  |  |  |  |  |
| Number recognition | .14 | -.15 | -.16 |  |  |  |  |  |  |
| Figure recognition | .22\* | .04 | -.02 |  |  |  |  |  |  |
| Object-number | -.21\* | .09 | .27\* |  |  |  |  |  |  |
| Number-figure | -.04 | -.11 | .18 |  |  |  |  |  |  |
| Figure-word | .02 | .18 | .26\* |  |  |  |  |  |  |
|  |  | Latent Variable Correlations | | | | | | | |
| Spatial | - |  |  |  |  |  |  |  |  |
| Verbal |  | - |  |  |  |  |  |  |  |
| Speed |  |  | - |  |  |  |  |  |  |
| Memory |  |  |  | - |  |  |  |  |  |

1. The horseshoe and horseshoe+ distributions lack closed form representations but are tightly bounded by definite upper and lower bounds. The densities plotted here represent the upper bound on the distributions as derived in Carvalho, Polson and Scott (2010) and Bhadra, Datta, Polson and Willard (2015). [↑](#footnote-ref-1)