Structural Equation Modeling: A Multidisciplinary Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

On the Performance of Horseshoe Priors for Inducing Sparsity in Structural Equation Models

Kjorte Harra & David Kaplan

To cite this article: Kjorte Harra & David Kaplan (19 Dec 2023): On the Performance of Horseshoe Priors for Inducing Sparsity in Structural Equation Models, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2023.2280895

To link to this article: https://doi.org/10.1080/10705511.2023.2280895

Published online: 19 Dec 2023.

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2023.2280895

On the Performance of Horseshoe Priors for Inducing Sparsity in Structural Equation Models

Kjorte Harra and David Kaplan

University of Wisconsin-Madison

ABSTRACT

The present work focuses on the performance of two types of shrinkage priors—the horseshoe prior and the recently developed regularized horseshoe prior—in the context of inducing sparsity in path analysis and growth curve models. Prior research has shown that these horseshoe priors induce sparsity by at least as much as the "gold standard" spike-and-slab prior. The horseshoe priors are compared to the ridge prior and lasso prior, as well as default non-informative priors, in terms of the percent shrinkage in the model parameters and out-of-sample predictive performance. Empirical studies using data from two large-scale educational assessments reveal the clear advantages of the horseshoe priors in terms of both shrinkage and predictive performance. Simulation studies reveal clear advantages in terms of shrinkage, but less obvious advantages in terms of predictive performance, except in the small sample size condition where both horseshoe priors provide noticeably improved predictive performance.

KEYWORDS

Bayesian regularization; horseshoe priors; prediction; structural equation modeling

1. Introduction

The development and implementation of Bayesian statistics in the social sciences have gained popularity in recent years with the increased accessibility of both proprietary and open-source software programs. As opposed to more traditional frequentist approaches, Bayesian methods arguably offer more intuitive interpretations and allow researchers to directly account for uncertainty in model parameters, variable selection via regularization, and model uncertainty (Kaplan, 2023).

This paper focuses on the flexibility offered with Bayesian inference in the context of variable section via model regularization. Model regularization enables researchers to induce sparsity and hence greater simplicity in their models without negatively impacting predictive performance. In other words, regularization directly addresses the age-old problem in statistics of the bias-variance trade-off—namely, the trade-off between the bias that can be incurred from not including potentially relevant variables in a model, with that of the loss in predictive accuracy due to increased variability across samples when variables have very small effects.

Regularization originated from a frequentist paradigm, specifically with the introduction of ridge regression by A. E. Hoerl and Kennard (1970), in which a penalty term is added to the regression equation to penalize the inclusion of irrelevant variables. However, more recent developments have demonstrated the flexibility and effectiveness of Bayesian approaches to regularization across a variety of statistical methods such as linear regression and factor analysis (see, e.g., Bainter et al., 2023; Jacobucci & Grimm, 2018; van Erp et al., 2019). As pointed out by van Erp (2020) there are a number of important benefits in adopting a Bayesian framework for model regularization and variable selection. First, as we will see, regularization can be easily implemented through the priors placed on model parameters, and these are generically referred to as *shrinkage* priors or *sparsity-inducing* priors. Shrinkage priors can be specified to shrink small coefficients toward zero while allowing large coefficients to remain large. Sparsity is induced by specifying certain hyperparameters within the priors specified for the model parameters. The hyperpriors can be manipulated to increase or decrease the amount of shrinkage in the estimated effects. For some regularization methods, these hyperparameters can be further defined through their own hyperprior distributions. Thus, model regularization fits the general framework of hierarchical Bayesian modeling.

The second benefit of adopting a Bayesian perspective to regularization is that the penalty term is estimated in the same step as the other model parameters. In other words, the penalty term is built into the model estimation process because it is incorporated directly into the model via a prior. In turn, that prior can be specified in a flexible manner through different settings, controlling for the degree of shrinkage as the researcher sees fit.

Finally, the third benefit of estimating Bayesian penalty terms via prior distributions is that many different forms of penalties can be implemented. There are frequentist-based penalty techniques, such as the ridge and lasso methods, which have their Bayesian counterparts, but in addition, there are methods that are strictly Bayesian such as the

CONTACT Kjorte Harra harra@wisc.edu Educational Psychology, University of Wisconsin-Madison, Madison, WI, USA.

spike-and-slab prior and the horseshoe priors which are the focus of this paper (see van de Schoot et al., 2021, for more information on these priors).

The present paper examines the performance of two relatively recent methods of Bayesian regularization: the original horseshoe prior (Carvalho et al., 2009, 2010a) and the regularized horseshoe prior (Piironen & Vehtari, 2017) in the context of path analysis and growth curve modeling. We compare these two horseshoe priors to the ridge prior and lasso prior in terms of the amount of shrinkage, but also in terms of out-of-sample predictive performance. Our focus on these two priors specifically is that they have been shown to perform at least as well as the "gold standard" spike-and-slab prior (Mitchell & Beauchamp, 1988) with the regularized horseshoe prior offering additional regularization of even large parameters. A further contribution of this paper is that we show how to implement the original and regularized horseshoe priors in the R package blavaan (Merkle & Rosseel, 2018).

The organization of this paper is as follows. In the next section, we introduce the general idea behind Bayesian regularization and then provide the details of the four regularization priors that will form the focus of this study. This is followed by a discussion of how we will assess the predictive performance of these priors. Then, we will demonstrate these priors in the context of two special cases of structural equation modeling; path analysis and growth curve modeling. For the path analysis example, the focus will be on a model of reading performance using data from the 2009 United States sample of the Program for International Student Assessment (PISA). For the growth curve modeling example, the focus will be on the prediction of change over time in reading performance using state-level data from the National Assessment of Educational Progress (NAEP). This is then followed by the design of our simulation study along with the results. The paper closes with a summary and discussion.

2. Overview of Bayesian Regularization

As noted earlier, Bayesian regularization differs from classical frequentist regularization in that frequentist regularization attaches a penalty term to a model parameter, whereas Bayesian regularization requires attaching a prior distribution to model parameters that promote penalizing small coefficients (Jacobucci & Grimm, 2018). Within the realm of Bayesian regularization, researchers have many prior distributions to choose from, beginning with the ridge prior (Hsiang, 1975) that seeks to shrink parameters close to zero and minimize collinearity. The Bayesian lasso, first introduced by Mitchell and Beauchamp (1988) improves upon the ridge prior as it allows shrinkage of coefficients to zero and thus serves as a method for variable selection. An extension of the frequentist lasso is the *adaptive lasso* which handles situations where the lasso is inconsistent for variable selection (Zou, 2006). We will not be exploring the adaptive lasso in this paper.

The Bayesian ridge and lasso priors, described in more detail below, are extensions of frequentist methods to the Bayesian context. Several strictly Bayesian approaches to regularization have also been developed. These include the *spike-and-slab* prior developed by Mitchell and Beauchamp (1988) and George and McCulloch (1993) which introduces a discrete prior to guide variable selection and sparsity. Next, the horseshoe prior by Carvalho et al. (2009, 2010a) allows for greater shrinkage than the ridge and the lasso while maintaining unregularized large coefficients. Most recently, to prevent large coefficients from remaining too large and escaping shrinkage, the regularized horseshoe was developed by Piironen and Vehtari (2017). The regularized horseshoe, which is sometimes referred to as the *Finnish Horseshoe*, also allows further flexibility than the original horseshoe prior and has been shown to further improve model predictive performance (see Piironen & Vehtari, 2017). A feature of the regularized horseshoe prior, as expanded on below, is that the user is required to specify the number of large parameters they believe are in the model.

In general, previous research has shown that Bayesian regularization can perform as well as, if not better than, classical methods of regularization in linear regression (van Erp et al., 2019). This finding has not been extended to path analysis nor growth curve modeling, and moreover, to our knowledge, has not been evaluated in terms of out-of-sample predictive performance. Thus, this paper focuses on the performance of two versions of the horseshoe prior: the original horseshoe prior and the regularized horseshoe prior, both described below, in the context of manifest variable path analysis and growth curve modeling as implemented in the R program blavaan (Merkle & Rosseel, 2018). By way of the empirical examples for each model type and a comprehensive simulation study for both, the original horseshoe and regularized horseshoe will be compared to two other popular regularization priors: the ridge and lasso prior, and to the case of no regularization in terms of the amount of shrinkage in path coefficients as well as out-of-sample predictive performance.

2.1. An Aside: The Spike-and-Slab Prior

For this paper, we will not be examining the *spike-and-slab* prior even though it has been considered the "gold-standard" for sparsity for quite some time (George & McCulloch, 1993; Mitchell & Beauchamp, 1988). However, in the interest of completeness, we should say a brief word about it.

Following Kaplan (2023), the spike-and-slab prior gets its name because the prior distribution on the individual regression coefficients come from a two-component mixture of Gaussian distributions and can be written as

$$\beta_p|\lambda_p, c, \epsilon \sim N(0, c^2) + (1-\lambda_p)N(0, \epsilon^2), \tag{1a}$$

$$\lambda_p \sim \text{Ber}(\pi) \tag{1b}$$

where $\lambda_p \in \{0, 1\}$ is an indicator variable that follows a Bernoulli distribution, denoted as Ber(·) and determines whether the coefficient is close to zero, in which case it

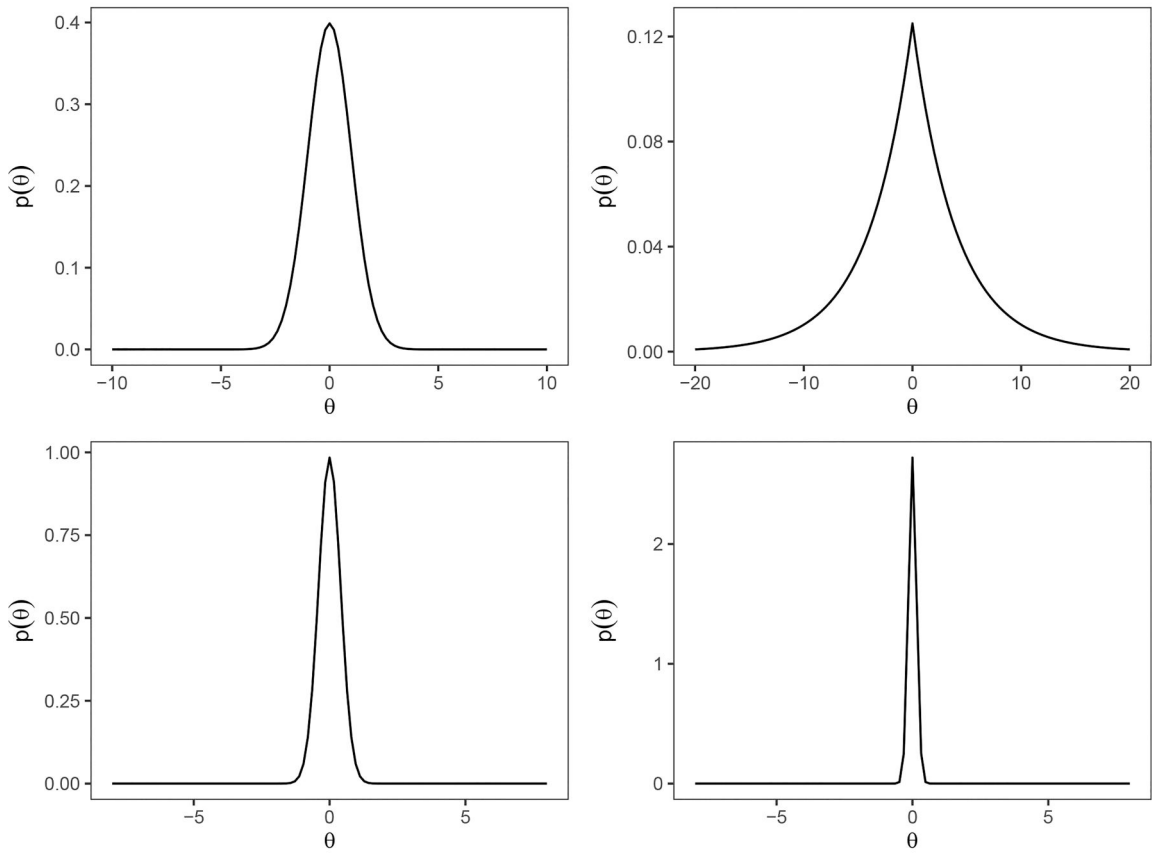

Figure 1. Regularization priors used in this paper. Top left: Ridge normal prior $N(0,1)$; top right: Lasso Laplace prior with location = 0, scale = 4; bottom left: Horseshoe prior with $\lambda_p \sim \mathcal{C}^+(0,1)$ and $\tau \sim \mathcal{C}^+(0,1)$; bottom right: regularized horseshoe prior with $\beta_j|\lambda_j,\tau,c \sim \mathrm{N}(0,\tau^2\tilde{\lambda}_j^2)$, where $\tilde{\lambda}_j^2 = \frac{c^2\lambda_j^2}{c^2+\tau^2\lambda_j^2}$, and $\lambda_j \sim \mathcal{C}^+(0,1)$.

comes from a so-called *Dirac delta* spike ($\lambda_p = 0$) or non-zero, in which case it comes from the slab ($\lambda_p = 1$) To create the spike, it is common to set $\epsilon = 0$. The slab width c and the inclusion probability π of the Bernoulli random variable is set by the user. Notice that with $\epsilon = 0$ the spike and slab prior can be rewritten as

$$\beta_p|\lambda_p, c \sim N(0, c^2\lambda^2), \tag{2a}$$

$$\lambda_p \sim \mathrm{Ber}(\pi). \tag{2b}$$

The result of this setup is that λ is a discrete parameter that only takes on two values ($\lambda_p = 0, 1$).

It may be interesting to note that the spike-and-slab prior with a Dirac delta spike at zero arguably violates *Cromwell's rule* (Lindley, 2007), which states that probabilities cannot be exactly zero (or one), insofar as it leads through Bayes' theorem to state complete uncertainty (or certainty) about a parameter, and thus does not allow for evidence to change one's assumptions about parameters via the posterior distribution (see also Kaplan, 2023).

Finally, it is necessary to point out that blavaan, which runs Stan in the background, cannot incorporate discrete parameters. However, studies have shown the similarity in performance between the spike-and-slab prior and the horseshoe prior (see, e.g., Carvalho et al., 2010b; Polson & Scott, 2011). Finally, the spike-and-slab prior is similar to the regularized horseshoe prior when the slab width $c < \infty$, thus providing some regularization on large coefficients. Thus, we prefer to study the horseshoe priors and demonstrate their implementation in blavaan.

2.2. Priors to Be Investigated

Figure 1 shows the density plots for the four regularization priors that we will be studying in this paper. We will refer to this plot throughout this section.

2.2.1. The Ridge Prior

As a regularization method, ridge regression (A. E. Hoerl & Kennard, 1970; R. W. Hoerl, 1985) aims to yield a parsimonious regularized regression model in the presence of highly correlated variables. Following the discussion in Kaplan (2023), the frequentist ridge estimator of $\boldsymbol{\beta}$, denoted as $\boldsymbol{\beta}_{ridge}$ is obtained by solving the minimization

$$\boldsymbol{\beta}_{ridge} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\boldsymbol{x}'\boldsymbol{x}) + \lambda\sum_{j=1}^{P}\beta_j^2, \tag{3}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of outcome scores, $\boldsymbol{\beta}$ is a $P \times 1$ vector of regression coefficients, $\boldsymbol{x}$ is an $n \times P$ matrix of predictors. The scalar $\lambda \geq 0$ is a tuning parameter that controls the degree of regularization and $\lambda\sum_{j=1}^{P}\beta_j^2$ is referred to as an $L_2 - norm$. When $\lambda = 0$, we have ordinary least squares, and when $\lambda = \infty$ we obtain $\boldsymbol{\beta}_{ridge} = \mathbf{0}$. With ridge

regression it can be seen that a large value of λ can lead to very heavy penalization.

The Bayesian specification of ridge regression was suggested by Hsiang (1975) who showed that if $\boldsymbol{\beta}$ has a mean of zero and covariance matrix $\boldsymbol{\Sigma} = (\sigma^2/\lambda)\mathbf{I}$, and if $\boldsymbol{\epsilon} \sim N(0, \sigma_{\epsilon}^2\mathbf{I})$, then the posterior mean of $\boldsymbol{\beta}$ is $(\boldsymbol{x}'\boldsymbol{x} + \lambda\mathbf{I})^{-1}\boldsymbol{x}'\boldsymbol{y}$, which is an alternative specification of the ridge estimator. Hsiang (1975) also notes that if weakly informative or informative priors are placed on each β_j, then the interpretation of the posterior mean of β as the ridge estimate is no longer valid. The penalty term (λ) is captured through normally distributed independent priors placed on the regression slope parameters. These normal priors have mean hyperparameter values fixed at zero in order to control shrinkage toward zero. The variance hyperparameter is typically rescaled to be in standard deviation form and is set to define the degree of spread that the distribution exhibits. Note that we specify a half-Cauchy prior distribution, denoted as $\mathcal{C}^+(0,1)$, for the residual standard deviation but other conjugate priors could be specified as well. A representation of the ridge prior is given in the top left of Figure 1.

2.2.2. The Lasso Prior

A drawback of ridge regression is that it does not improve parsimony in the sense that all of the variables still remain in the model after penalization (Zou & Hastie, 2005). A method that appears similar to ridge regression but is principally different in terms of yielding a parsimonious model is the *least absolute shrinkage and selection operator* (Tibshirani, 1996).

The frequentist lasso involves solving the expression

$$\boldsymbol{\beta}_{lasso} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\boldsymbol{x}'\boldsymbol{x}) + \lambda\sum_{j=1}^{P}|\beta_p|, \tag{4}$$

The term $\lambda\sum_{j=1}^{P}|\beta_j|$ is referred to as an $L_1 - norm$ penalty which allows less important coefficients to be set to zero, and thus the lasso provides for both shrinkage and variable selection. The Bayesian lasso, first introduced by Park and Casella (2008), uses a double exponential or Laplace prior where

$$p(\beta_j) = \frac{1}{2\tau}\exp\left(-\frac{|\beta_j|}{\tau}\right), \tag{5}$$

where $\tau = 1/\lambda$.

The top right of Figure 1 shows the double exponential distribution. We see that the double exponential distribution is ideal because it peaks at zero, which shrinks small coefficients toward zero. However, the double exponential distribution can be set to have thick tails (in both directions), allowing the larger coefficients to remain large. Given that the distribution is centered at zero to control shrinkage toward zero, the mean hyperparameter setting is fixed to zero. The scale, or dispersion, of the double exponential distribution, is the hyperparameter that researchers can alter when implementing the lasso. This defines the amount of spread and the thickness of the tails, which controls the degree of shrinkage in coefficients. Again, a $\mathcal{C}^+(0,1)$ prior can be specified on the standard deviation of the residuals, if desired.

Although the ridge and lasso priors are similarly implemented in the Bayesian framework, these techniques can produce different amounts of shrinkage depending on the hyperparameter settings. That is, the lasso approach can result in more shrinkage for the small estimates, but less shrinkage for the large estimates. This result is a function of the double exponential distribution implemented in the lasso approach. The double exponential distribution is more peaked around zero and it has heavier tails compared to the normal distribution used in the ridge approach. Regardless of the approach implemented, Bayesian penalization can be a useful tool when attempting to avoid overfitting a complex model to small samples. Indeed, the lasso is simultaneously a shrinkage and variable selection method. In addition, these approaches further highlight the modeling flexibility that Bayesian methods provide through the flexible implementation of priors.

2.2.3. Two Horseshoe Priors

The focus of this paper is on the performance of two horseshoe priors: the original horseshoe prior first introduced by Carvalho et al. (2009, 2010a), and the more recently developed regularized horseshoe prior introduced by Piironen and Vehtari (2017).

2.2.3.1. Original Horseshoe. The horseshoe prior can be characterized as a scale mixture of normals with half-Cauchy tails offering unique features in enacting shrinkage not previously seen in other regularization priors. Specifically, the horseshoe prior is defined as follows. For $j = 1, \ldots, P$

$$\theta_j|\lambda_j, \tau \sim \mathcal{N}(0, \lambda_j^2\tau^2) \tag{6a}$$

$$\lambda_j \sim \mathcal{C}^+(0,1), \tag{6b}$$

where τ is the global shrinkage parameter on model parameters θ_j, and λ_j is the local shrinkage parameter. A unique feature of the horseshoe prior is that the tails of the $\mathcal{C}^+$ distribution permit large parameters to remain unregularized, while the global shrinkage parameter τ severely shrinks parameters that are close to zero (Carvalho et al., 2009, 2010a). The density plot for the original horseshoe is given in the lower left of Figure 1.

Gaining an intuition of the horseshoe prior can be obtained by studying its so-called *shrinkage profile*. Following Piironen and Vehtari (2017), first assume that the predictors, $\{x_j\}_{j=1}^{p}$ are uncorrelated with mean zero and variances $\mathbb{V}(x_j) = s_j^2$. Further, let $\hat{\beta}_j$ be the maximum likelihood estimate of β_j. Then, the shrunken estimate of β_j can be approximated by

$$\bar{\beta}_j = (1 - \kappa_j)\hat{\beta}_j, \tag{7}$$

where κ_j is the shrinkage factor for β_j defined as

$$\kappa_j = \frac{1}{1 + n\sigma^{-2}\tau^2 s_j^2\lambda_j^2} \tag{8}$$

We see that as $\tau \to \infty, \overline{\beta}_j \to \hat{\beta}_j$, implying no shrinkage. Conversely, as $\tau \to 0, \overline{\beta}_j \to 0$, implying total shrinkage.

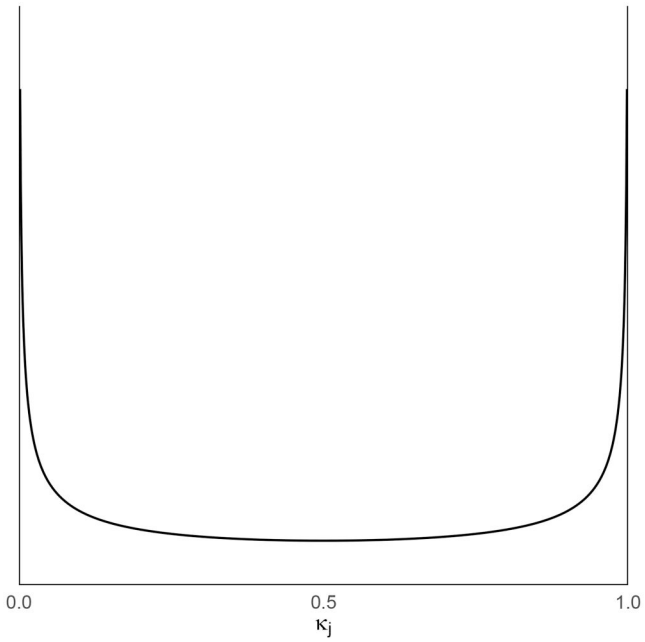

Figure 2. Density of shrinkage weight for the original horseshoe prior. Note that the profile suggests high density when there are large parameters ($\kappa \approx 0$, no shrinkage) and high density when there are very small parameters ($\kappa \approx 1$, total shrinkage) (Carvalho et al., 2010a).

Furthermore, Carvalho et al. (2010a, see also; Piironen & Vehtari, 2017) show that when λ_j take on independent half-Cauchy distributions, then κ_j takes on the horseshoe-shaped Beta $\left(\frac{1}{2}, \frac{1}{2}\right)$ distribution from which the prior gets its name (see Figure 2). We observe that when small parameters are encountered, there is a high probability of shrinkage, and when there are large parameters encountered, there is a high probability of no shrinkage.

An added benefit of the horseshoe prior over the ridge and lasso priors is that it can distinguish global shrinkage using τ and local shrinkage with λ_j (Carvalho et al., 2010a). Furthermore, the horseshoe prior is essentially a continuous counterpart to the spike-and-slab prior with an infinitely wide slab width (Piironen & Vehtari, 2017) and thus offers a method of regularizing models in a similar manner as spike-and-slab priors when using software that does not support such discrete parameter distributions, such as Stan (Stan Development Team, 2021) and programs that rely on Stan such as rstanarm (Goodrich et al., 2022), brms (Bürkner, 2017, 2018), and blavaan (Merkle et al., 2021; Merkle & Rosseel, 2018).

2.2.3.2. Regularized Horseshoe. Following the discussion given in Piironen and Vehtari (2017, see also; Kaplan, 2023), a limitation of the original horseshoe prior relates to the regularization of the large coefficients. Specifically, it is still the case that large coefficients can transcend the global scale set by τ_0 with the impact being that the posteriors of these large coefficients can become quite diffused, particularly in the case of weakly-identified coefficients (Betancourt, 2018a). To remedy this issue, Piironen and Vehtari (2017) proposed a *regularized* version of the horseshoe prior (also known as the *Finnish horseshoe prior*). Following the notation used in Betancourt (2018a) the regularized horseshoe prior takes the form of the following: For $j = 1, \ldots, p$, where p are the number of predictors,

$$\beta_j \sim \mathcal{N}(0, \tau^2 \tilde{\lambda}_j^2), \tag{9a}$$

$$\tilde{\lambda}_j = \frac{c\lambda_j}{\sqrt{c^2 + \tau^2 \lambda_j^2}}, \tag{9b}$$

$$\lambda_j \sim \mathcal{C}^+(0, 1), \tag{9c}$$

$$c^2 \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2} s^2\right), \tag{9d}$$

$$\tau \sim \mathcal{C}^+(0, \tau_0), \tag{9e}$$

where $c > 0$ and s^2 is the variance for each of the p predictor variables. As pointed out by Piironen and Vehtari (2017), those variables that have large variances would be considered more relevant a priori, and while it is possible to provide predictor-specific values for s^2, generally we scale the variables ahead of time so that $s^2 = 1$. Finally, c^2 is the slab width which controls the size of the large regression coefficients. The density plot for the regularized horseshoe is given in the lower right of Figure 1.

To gain an intuition of the regularized horseshoe, first note that the form of Equations (9a-9d) is quite similar to the horseshoe prior, however $\tilde{\lambda}_j$ places a control on the size of the coefficients by introducing a slab width c^2 in Equation (9b). Following Piironen and Vehtari (2017), notice that if $\tau^2\lambda_j^2 \ll c^2$ then this means that β_j is close to zero and $\tilde{\lambda}_j \to \lambda_j$, which is the original horseshoe in Equations (6a) and (6b). However, if $\tau^2\lambda_j^2 \gg c^2$, then $\tilde{\lambda}_j \to c^2/\tau^2$ and the prior begins to approach the $\mathcal{N}(0, c^2)$, where again, the choice of c^2 controls the size of the large coefficients. Because c^2 is a slab width that might not be well known, it follows that it should be given a prior distribution, and Piironen and Vehtari (2017) recommend the inverse-gamma distribution ($\mathcal{IG}$) in Equation (9d) which induces a relatively non-informative Student's—t slab when coefficients are far from zero (Piironen & Vehtari, 2017). When coefficients are small, the prior behaves similarly to the horseshoe, but when coefficients are large they are regularized with a Gaussian slab with a variance of c^2.

Another way to understand the regularized horseshoe prior is to examine its shrinkage profile relative to the shrinkage profile of the original horseshoe in Figure 3. Below, we present the shrinkage profile of the regularized horseshoe. Note that this shrinkage profile does not have the exact shape of the Beta $\left(\frac{1}{2}, \frac{1}{2}\right)$ distribution of the original horseshoe in Figure 3. Specifically, there is slightly more mass at the left mode because the regularized horseshoe is shifted to the right of 0.0 by 0.05 (see also Piironen & Vehtari, 2017). Thus, the regularized horseshoe induces a certain amount of shrinkage ($\kappa_j = 0.05$) when encountering large parameters and so they too are regularized as are the small parameters under total shrinkage ($\kappa_j = 1.0$).

2.2.3.3. Eliciting Hyperparameters for τ. Inspection of Equation (9e) reveals that one needs to set a value of τ_0 which, in turn, will determine the amount of global shrinkage induced by τ in Equation (9a). The approach we study

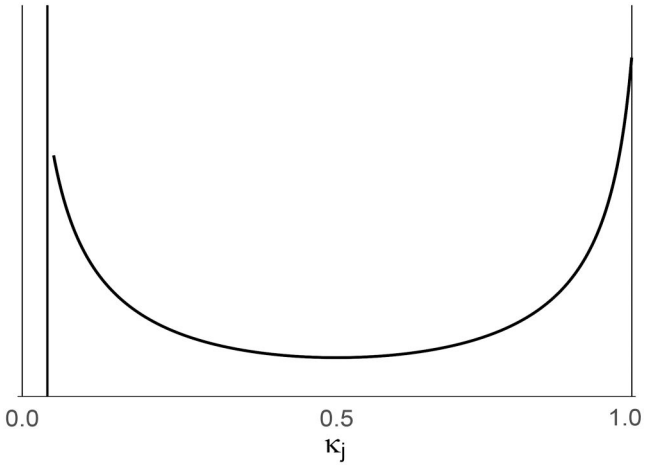

Figure 3. Density of shrinkage weight for the regularized horseshoe prior. Note that this shrinkage profile does not have the exact shape as the original horseshoe in Figure 1 there is slightly more mass at the left mode (see also Piironen & Vehtari, 2017).

in this paper rests on the intuition provided in Piironen and Vehtari (2017) who relate the choice of τ_0 to the effective number of non-zero coefficients in the model. Specifically, Piironen and Vehtari (2017) define the effective number of non-zero coefficients in terms of the shrinkage parameter κ_j as

$$m_{eff} = \sum_{j=1}^{p}(1-\kappa_j) \tag{10}$$

where we can see that if κ_j is close to zero or one, the result is effectively the number of non-zero coefficients in the model. Next, Piironen and Vehtari (2017) show that assuming that $\mathbb{V}(x_j) = s_j = 1$, then

$$\mathbb{E}(m_{eff}|\tau,\sigma) = \frac{\tau\sigma^{-1}\sqrt{n}}{1+\tau\sigma^{-1}\sqrt{n}} \times p \tag{11}$$

Now, letting $p_0 \equiv \mathbb{E}(m_{eff}|\tau,\sigma)$, a value of τ_0 for τ obtains as

$$\tau_0 = \frac{p_0}{p-p_0} * \frac{\sigma}{\sqrt{n}} \tag{12}$$

where again, p is the number of coefficients in the model, and p_0 $(\neq p)$ is the number of coefficients assumed to be "large" in the model.

For practical purposes, Piironen and Vehtari (2017) note that the standard choice of a $\mathcal{C}^+(0,1)$ distribution for τ would end up placing a great deal of mass on large coefficients, particularly when p is large, and thus the solution would end up with many coefficients that are "unshrunk." A better approach is to use prior information to decide on a sensible number for p_0 with the understanding that because the contexts of research questions can change quite dramatically, there are no global rules of thumb for choosing values of p_0. It should also be noted that previous research has demonstrated that the regularized horseshoe prior is robust to the choice of p_0 (Piironen & Vehtari, 2017), but to the best of our knowledge, this has not been replicated in the context of structural equation models (though we don't expect many differences in this context) nor, more importantly, with respect to out-of-sample prediction as will be studied in this paper. Thus, a contribution of this paper will be to examine the performance of the regularized horseshoe under different values of p_0 through our simulation study.

3. Evaluating Predictive Performance

Arguably, the overarching goal of statistics is prediction. In other words, a key characteristic of statistics is to develop accurate predictive models, and all other things being equal, a given model is to be preferred over other competing models if it provides better predictions of what actually occurred (Dawid, 1982, 1984, see also; Kaplan, 2023). Indeed, it is hard to feel confident about inferences drawn from a model that does a poor job of predicting the extant data. The problem, however, is how to develop accurate predictive models, and, importantly, how to evaluate their accuracy. Thus, instead of attempting to simply explain the current data at hand, a model should be applicable to other data outside the present sample.

For this paper, we will be using Bayesian leave-one-out cross-validation (LOO-CV) to evaluate model predictive performance (Vehtari et al., 2017). A unique contribution of this work is using LOO-CV as a method of comparing the predictive performance of the four regularization priors. Bayesian leave-one-out cross-validation is a special case of so-called *k-fold* cross-validation in which the data set is divided into two folds: The statistical model of interest is fitted with the training set and then compared to the i^{th} observation in the test set to measure predictive performance. This process is then repeated for each observation being designated as the "left out" data point. Thus, LOO-CV is a k-fold cross-validation procedure where $k = n$.

The LOO-CV has been around since at least the mid-1970s (Allen, 1974; Stone, 1974) and is uniquely suited to the question of out-of-sample predictive performance. It is superior to the deviance information criterion (DIC) (Spiegelhalter et al., 2002) insofar as the DIC, like its frequentist counterpart the *Akaike information criterion* (AIC) (Akaike, 1973) is not fully Bayesian because it conditions on a plug-in point estimate, say $\hat{\theta}_{Bayes}$, (albeit Bayesian) rather than averaging over the posterior distribution to account for uncertainty in the parameter estimates. As noted by Gelman et al. (2014), the AIC and DIC are actually gauging the predictive performance of the plug-in estimators, say, $\hat{\theta}_{mle}$ and $\hat{\theta}_{Bayes}$, respectively, rather than the actual predictive density. What is needed therefore is a measure based on the pointwise predictive density for each observation in the data set as these are the quantities of actual concern in prediction.

The LOO-CV is quite similar to the *widely applicable information criterion* (WAIC) was developed by Watanabe (2010) as a fully Bayesian counterpart to the AIC. As with the AIC and DIC, the WAIC uses the same data to assess the predictive performance of the model as was used to estimate the model and thus could lead to overstating the predictive quality of the model. Although the WAIC and LOO-CV are

asymptotically equivalent, ideally, we would like to use actual out-of-sample data in order to assess predictive performance.

The LOO-CV rests on the derivation of the *expected log point-wise predictive density* (ELPD) for new data defined as

$$\text{ELPD} = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log\ p(\tilde{y}_i|y) d\tilde{y}_i, \tag{13}$$

where $p_t(\tilde{y}_i)$ represents the distribution of the true but unknown data-generating process for the predicted values $\tilde{y}_i$ and where Equation (13) is approximated by cross-validation procedures. The ELPD provides a measure of predictive accuracy for the n data points taken one at a time (Vehtari et al., 2017).

The ELPD for leave-one-out cross-validation (denoted as ELPD_{loo} is defined as

$$\text{ELPD}_{loo} = \sum_{i=1}^{n} \log\ p(y_i|y_{-i}), \tag{14}$$

where

$$p(y_i|y_{-i}) = \int p(y_i|\theta) p(\theta|y_{-i}) d\theta \tag{15}$$

is the LOO predictive density given the data with the i^{th} data point left out (Vehtari et al., 2017). The log sum of these predictive densities in Equation (14) is the LOO-CV estimate of the expected log pointwise predictive density (ELPD) (Gelman et al., 2014; Gronau & Wagenmakers, 2019; Vehtari et al., 2017).

As pointed out by Vehtari et al. (2017), LOO is asymptotically equivalent to the WAIC, but in the case of finite samples with weak priors and/or influential observations, a more robust method for calculating the LOO-CV might be desired. To this end, Vehtari et al. (2017) developed a fast and stable approach to obtaining LOO-CV using Pareto-smoothed importance sampling (PSIS). As applied to LOO, the importance ratios are obtained as

$$r_i^t = \frac{p(\theta_t|y_{-i})}{p(\theta^t|y)} \tag{16}$$

From here, the importance sampling LOO predictive distribution is obtained as

$$p(\tilde{y}_i|y_{-i}) \approx \frac{\sum_{t=1}^{T} r_i^t\ p(\tilde{y}_i|\theta^t)}{\sum_{t=1}^{T} r_i^t}. \tag{17}$$

The density of the held-out data point is obtained from the T posterior samples as

$$p(y_i|y_{-i}) \approx \frac{1}{\frac{1}{T}\sum_{t=1}^{T} \frac{1}{p(y_i|\theta^t)}}. \tag{18}$$

The importance weights in Equation (16) can have very large (and perhaps infinite) variances. To handle this, Vehtari et al. (2017) propose the use of the generalized Pareto distribution which provides the same diagnostic based on the shape parameter k of the Pareto distribution. The same rules of thumb as discussed in Section 4.8 apply. The PSIS approach is implemented in loo (Vehtari et al., 2019), and is often referred to as PSIS-LOO. The PSIS-LOO measure is available in the R software program loo (Vehtari et al., 2019) and available in blavaan (Merkle & Rosseel, 2018) when calling **"blavFitIndices()."**

The LOO measure offers point-wise predictions where individual observations can be predicted as opposed to overall means. Previous research has demonstrated that LOO-CV values are a sufficient way to measure a model's out-of-sample predictive fit (Vehtari et al., 2017). By comparing the same path model with different regularization priors one can directly compare regularization methods in terms of out-of-sample predictive performance.

It is useful to note that an information criterion based on LOO, referred to as the *LOO-IC*, can be easily derived as

$$\text{LOO} - \text{IC} = -2\ \widehat{\text{elpd}}_{loo} \tag{19}$$

which places the LOO-IC on the deviance scale. Among a set of competing models, the one with the smallest LOO-IC is considered best from an out-of-sample point-wise predictive point of view. We will use the LOO-IC for the comparison of our regularization priors. In addition, it may also be interesting to note that under maximum likelihood estimation, LOO-CV is asymptotically equivalent to the AIC (Stone, 1977, see also; Yao et al., 2018).

4. Empirical Examples

The following section will demonstrate how both versions of the horseshoe prior can be implemented and performance can be compared to other regularization priors using real-world data in structural equation modeling—specifically path analysis and growth curve modeling. For both empirical examples, R's blavaan (Merkle et al., 2021) package is used to estimate the models and calculate LOO-IC estimates to compare predictive performance.

4.1. Path Analysis

The empirical data used in the present path analysis example originates from the 2009 US sample of PISA ($N = 5{,}233$). The aim of PISA is to assess secondary students' competencies in several academic domains as well as students' motivations and learning strategies (OECD. , 2010). Participating 15-year-old students complete an exam with a focus on reading literacy for the 2009 cycle, as well as a background questionnaire to gain a better understanding of students' attitudes, motivations, and knowledge of the broader social contexts in which they learn. PISA is used to inform education policymakers of factors associated with successful educational outcomes as well as markers for countries' progress toward educational goals.

Figure 4 below displays this path model, where the outcome is reading comprehension scores as predicted by three approaches to students' learning: memorization through rereading texts until they can recite them (MEMOR), elaboration strategies such as relating the material to their prior knowledge (ELAB), and control strategies such as checking for understanding (CSTRAT). Predictors of these learning strategies include a PISA index of economic, social, and cultural status (ESCS), a binary indicator of students' gender (0 = male, 1 = female), and an index of immigrant

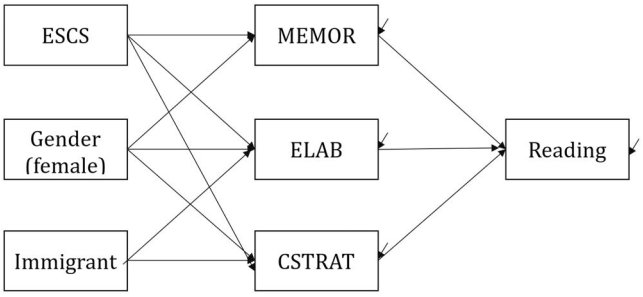

Figure 4. Path model predicting US PISA students' reading scores by three different learning strategies (memorization, elaboration, and control strategies) as well as student background information (economic, social, and cultural status, gender, and immigration background).

background (Immigrant) which asks if a student is a first-generation, second-generation, or native-born student.

4.1.1. Bayesian Path Analysis

The general form of a path analytic model among observable variables can be written in matrix notation as (see e.g Bollen, 1989; Kaplan, 2009)

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \tag{20}$$

where $\boldsymbol{\alpha}$ is a vector of structural intercepts, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ represent the structural relationships among the variables and $\boldsymbol{\zeta}$ is a vector of residuals. Estimation of this model from a frequentist perspective is based on principles of either maximum likelihood estimation or generalized least squares. Bayesian estimation of the path model requires that prior distributions be placed on all model parameters, and summaries of the posterior distributions of the model parameters are obtained via methods of Markov chain Monte Carlo sampling (MCMC). For a detailed discussion of Bayesian path analysis, and Bayesian structural equation modeling in general, see Depaoli (2021).

4.1.2. Blavaan Implementation

A contribution of this paper includes a guide on how researchers can implement both horseshoe priors into blavaan utilizing the **"mcmcextra = list()"** argument in blavaan functions such as **"bsem"** used in this paper. However, a few details are worth noting. The present version of blavaan runs the Stan program language in the background, and Stan utilizes *Hamiltonian Monte Carlo* (HMC) via the *No-U-Turn* (NUTS) sampler. For a discussion of HMC and NUTS, see Betancourt (2018b) and Hoffman and Gelman (2014). However, in order to implement the horseshoe priors in blavaan it is necessary to add code to blavaan via the **"mcmcextra = list()"** command. At present however, **"mcmcextra"** cannot be implemented if Stan is used, and so the analysis must be run using JAGS (Plummer, 2003) in the background via the **"target ="jags"."**[1] The full blavaan code for our empirical example, including the specification of the priors for each regularization method, is given in Appendix A.

4.1.3. Results of Empirical Example

Preliminary analyses revealed that very little differences were found for sample sizes larger than 100. Thus, for this empirical example, a path model was estimated on a random sample of size 50. Markov chain Monte Carlo sampling using the Gibbs sampler implemented in blavaan through its interface with JAGS was constructed with 10,000 iterations, 5,000 burn-in and 5,000 post-burn-in. The models took approximately three minutes each to compute.

Results of the path analysis with default non-informative priors along with the four different regularization priors used are depicted in Figure 5 with mean parameter estimates and their 95% probability intervals. Bayesian credible intervals represent the probability that the true effect of a predictor on an outcome lies within that interval. Model diagnostics including *potential scale reduction factors* (Brooks & Gelman, 1998) along with other MCMC convergences indices such as trace plots and autocorrelation plots (not shown) revealed adequate convergence of the algorithm.

Posterior regression estimates for the blavaan default non-informative priors and all of the regularization priors show roughly the same pattern of results, but considerable differences in the amount of shrinkage. For example, in Figure 5, the results under the blavaan default priors show that the use of elaboration strategies for reading is predicted by the PISA socioeconomic measure (ESCS) and that zero is not in the 95% credible interval. Both the ridge and lasso priors show a small (and similar) amount of shrinkage, but zero remains outside the 95% credible interval. Contrast this with the results for the horseshoe and regularized horseshoe priors where we find considerable (and comparable) shrinkage, but now zero resides in the 95% credible interval. We find no noticeable differences in parameter estimates between the horseshoe and regularized horseshoe, and both versions of the horseshoe yielded a greater amount of shrinkage as well as slightly more precise parameter estimates compared to the previously used priors and a non-informative unregularized default prior.

Figure 6 more clearly shows the amount of shrinkage based on each prior relative to the blavaan non-informative default prior model. The height of each bar segment represents the mean parameter estimate for each regularization prior. Segments smaller than the default prior estimates demonstrate more shrinkage for a given prior on a given parameter. We can see that the horseshoe and regularized horseshoe demonstrate the largest amount of shrinkage across parameters, especially those that are relatively large. We most strongly observe this effect with the "READ ~ CSTRAT," "ELAB ~ ESCS", and "CSTRAT ~ GENDER" parameters, as both versions of the horseshoe priors have noticeably smaller estimates than the ridge and lasso priors. We do not observe much difference in terms of shrinkage between the original and regularized horseshoe.

[1]Note that the spike-and-slab prior cannot be implemented in Stan because of the program's inability to implement discrete prior distributions (Stan Development Team, 2021).

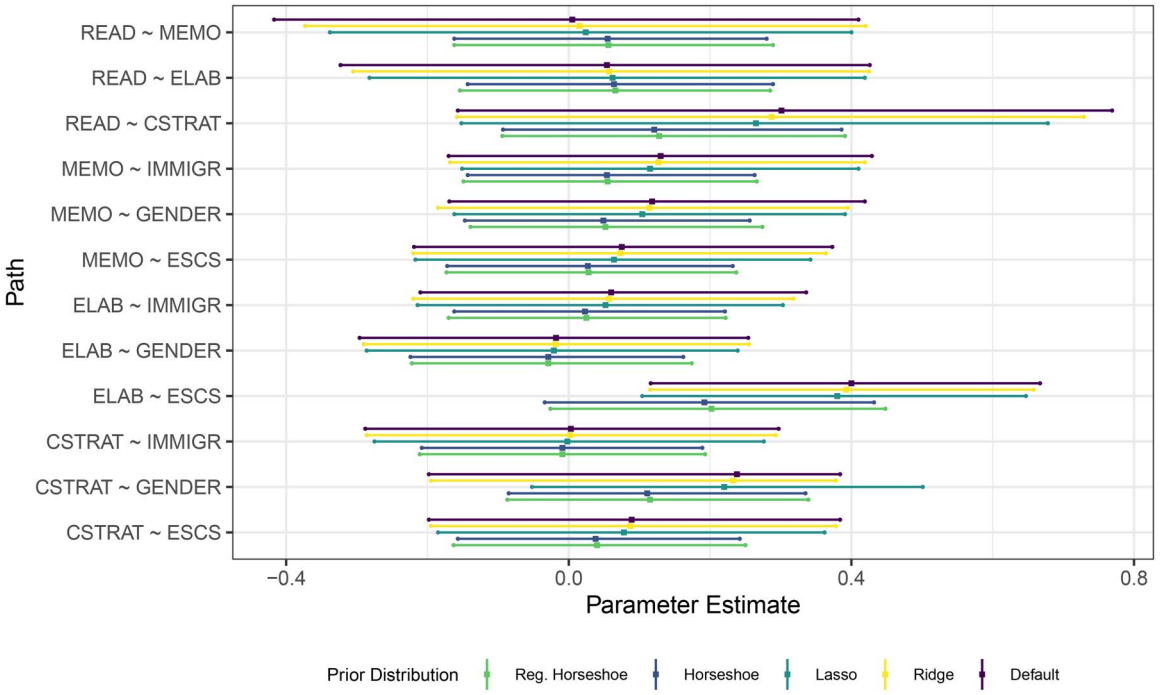

Figure 5. Empirical path model results for predicting US PISA students' reading achievement ($n = 50$).

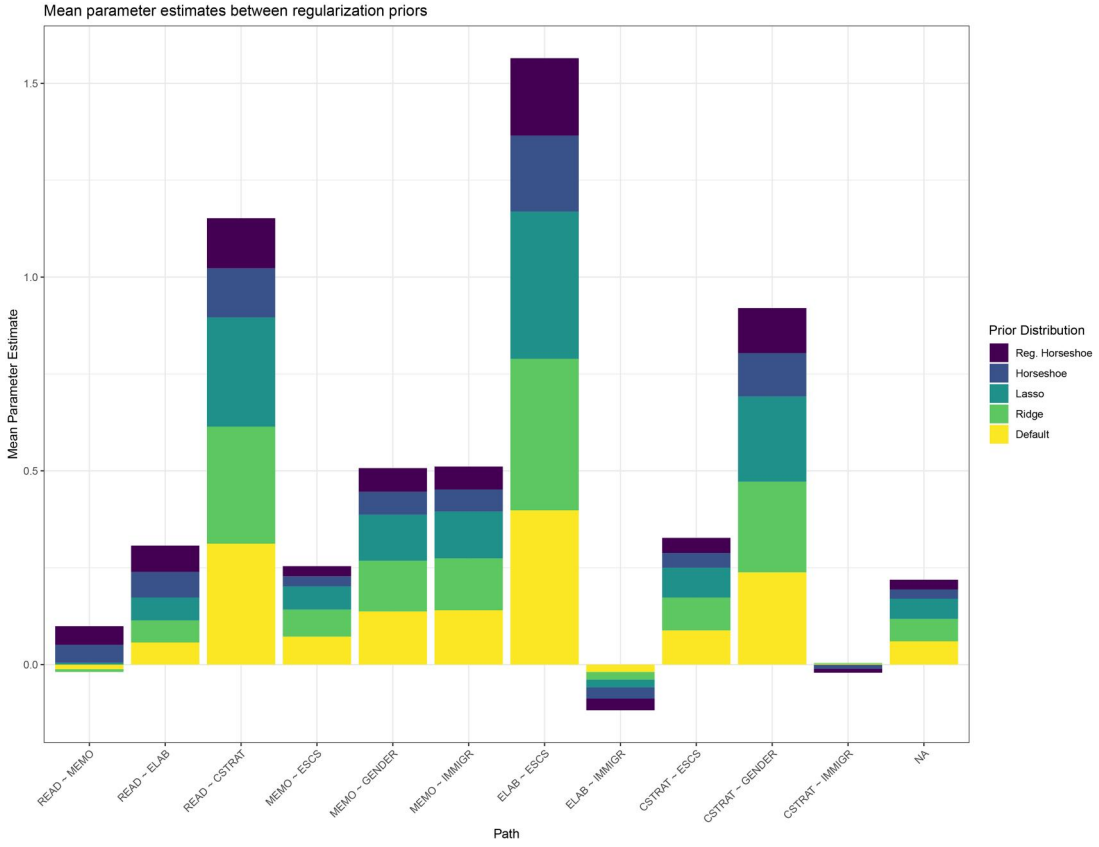

Figure 6. Mean parameter estimates for the given path model in the empirical study.

To assess the predictive performance of the four regularization priors, LOO-IC values were used to compare out-of-sample predictive performance. LOO-IC values for each of the four regularization priors and the blavaan default priors are reported in Table 1. Compared to other competing models, the model with the lowest LOO-IC value holds the strongest predictive accuracy.

We find that the horseshoe priors provide the best predictive performance compared to the ridge and lasso priors in predicting reading achievement. Default path analysis with non-informative priors and no regularization performed the worst, followed by the ridge and then the lasso. Based on these results, it is evident that the original

Table 1. LOO-IC estimates for each regularization method in the path analysis example.

	Estimate	Difference from default
Default	588.53	0.00
Ridge	587.42	−1.11
Lasso	585.32	−3.21
Horseshoe	579.11	−9.42
Regularized horseshoe ($p_0 = 1$)[a]	579.26	−9.27

[a]$p_0 = 1$ means that we are assuming that there is only one large coefficient in the model. Differences are calculated from blavaan default non-informative priors.

Table 2. LOO-IC estimates for each regularization method in the growth curve model.

	Estimate	Difference from default
Default	216.86 (20.82)	0.00
Ridge	217.07 (20.93)	−0.21
Lasso	216.82 (20.93)	−0.04
Horseshoe	212.32 (20.83)	−4.54
Regularized horseshoe ($p_0 = 1$)[a]	212.60 (20.82)	−4.26

[a]$p_0 = 1$ means that we are assuming that there is only one large coefficient in the model. Differences are calculated from blavaan default non-informative priors.

horseshoe and regularized horseshoe Bayesian priors are a reasonable choice in path analysis as they allow large coefficients to remain large while shrinking small ones while at the same time providing the overall best predictive performance. Other analyses with larger sample sizes showed very few differences and our results are in line with Jacobucci and Grimm (2018) and consistent with Bayesian theory that when sample sizes are large the prior distribution becomes less influential than the likelihood.

4.2. Growth Curve Modeling

The growth model we use for this empirical example includes data from eight cycles of the National Assessment of Educational Progress (NAEP) from 2003 to 2017 (U.S. Department of Education et al., 2022) as well as other data from the National School Lunch Program (NSLP) and the Common Core of Data (CCD, Common Core of Data, 2020). We construct a model tracking the rate of growth in students' math scores over time across US states, as predicted by reading scores at a given time point as well as several school-level time-invariant predictors. These variables include the difference between 2017 and 2003 state-level full-time equivalent teachers, the difference between 2017 and 2003 pupil-to-teacher ratio, the difference between 2015 and 2003 total revenue, and the difference in the 2017 and 2003 percentages of NSLP-eligible students. Specifically, we write the intra-state model of change over time in mathematics scores as (see Bollen & Curran, 2006; Depaoli, 2021)

$$y_{ti} = \pi_{0i} + \pi_{1i} a_{ti} + r_{ti} \tag{21}$$

where y_{it} is mathematics score for state i $(i = 1, \ldots n)$ at time t $(t = 1, \ldots T)$ where a is an indicator of the NAEP assessment cycle, π_{0i} is the intercept capturing state i's math score at time t, π_{1i} is the slope (rate of linear growth over time) in the mathematics score for state i at time t, and r_{it} is the residual term. Together π_0 and π_1 are the *growth parameters*.

The inter-state model can be written generally as

$$\pi_{pi} = \beta_{p0} + \sum_{k=1}^{K_p} \beta_{pk} x_{ki} + \epsilon_{pi}, \tag{22}$$

where the π_{pi} are the growth parameters (intercept and growth rate), x_{ki} are values on K predictors for state i, and ϵ_{pi} are errors.

Bayesian growth curve modeling requires specifying a probability model for the outcome and placing priors on all model parameters (Kaplan, 2023). The priors on this growth curve model are non-informative or weakly-informative (see e.g., Gelman et al., 2017).

A total sample size of 50 jurisdictions was used, including 49 US states (Tennessee was removed due to missingness) as well as Washington D.C. Markov chain Monte Carlo sampling using the Gibbs sampler implemented in blavaan's "**bgrowth"** function through its interface with JAGS was constructed with 20,000 iterations and 5,000 adaption steps. The models each took approximately fifteen minutes to compute. The full blavaan code for this empirical example, including the specification of the regularization priors, can be found in Appendix A.

4.2.1. Results

We find with this empirical example that the horseshoe priors again outperformed other regularization priors and a non-informative default prior. Table 2 reports these results, we see very little difference in terms of predictive performance between the ridge, lasso, and default priors. However, the lower LOO-IC estimates for both horseshoe priors provide evidence that those priors induced stronger predictive performance than the other methods.

Pairwise comparisons of each regularization method provide further evidence that in this example the horseshoe priors outperform the others. Table 3 contains pairwise comparisons of LOO estimates, where a negative difference means the regularization method in that row exhibits a lower LOO estimate. Following Hollenbach and Montgomery (2020)'s recommendations, which assert that a LOO difference at least twice its standard error is noteworthy, we find that both horseshoe priors possess stronger predictive performance than the ridge, lasso, and unregularized model. The regularized horseshoe also performed better than the original horseshoe when directly comparing distributions, despite the regularized horseshoe holding a slightly larger mean LOO-IC estimate as seen in Table 2.

These two empirical examples with growth curve modeling and path analysis suggest that the original horseshoe prior and regularized horseshoe prior both result in greater amounts of shrinkage and noticeably better predictive performance in comparison to the ridge and lasso priors for small sample sizes with real-world data.

5. Simulation Study

In this section, we present a comprehensive simulation study comparing the horseshoe and regularized horseshoe

Table 3. LOO pairwise differences and standard errors for each regularization method in the growth curve model.

	Default	Ridge	Lasso	Horseshoe
Default				
Ridge	−0.10 (0.09)			
Lasso	−0.05 (0.06)	**−0.14** (0.06)		
Horseshoe	**−2.29** (0.29)	**−2.39** (0.31)	**−2.24** (0.31)	
Regularized horseshoe	**−2.13** (0.30)	**−2.22** (0.32)	**−2.08** (0.32)	**−0.17** (0.04)

Differences ≥ 2 standard errors are bolded.

priors to the ridge and lasso priors with respect to the amount of shrinkage in parameter estimates as well as predictive performance, while varying the sample size and the number of coefficients that the investigator has reason to believe (a priori) are large, which is specific to the regularized horseshoe prior.

5.1. Path Analysis Simulation Study

5.1.1. Conditions

A Monte Carlo simulation study was conducted to compare the predictive performance and average shrinkage of these regularization priors in Bayesian path analysis. Data were simulated from the same path model previously used in our empirical example except the path of gender predicting elaboration was removed due to its insignificance. For blavaan defaults, the ridge prior, the lasso prior, and the horseshoe prior, four different sample size conditions were tested ($n = 30, 100, 500,$ and 3000). For the regularized horseshoe prior, the same sample size conditions were tested as well as the choice of p_0, where $p_0 = 1, 3, 6,$ and 10 where the total number of parameters p is 11. In total, there were 32 conditions. One thousand replications were generated for each condition of the study. For each replication, models were estimated with Markov chain Monte Carlo sampling using the Gibbs sampler in blavaan with 10,000 iterations, 5,000 burn-in, and 5,000 adaptation steps. Simulation results showed acceptable model convergence for further analysis. Software code for the simulation study is available at http://bmer.wceruw.org/publications.html.

5.1.2. Results of Simulation Study

Figure 7 below depicts mean parameter estimates across regularization priors and sample sizes for the given path model. For both the ridge and lasso priors, shrinkage is relatively small across parameters, especially when sample sizes are larger. The lasso does yield greater shrinkage than the ridge particularly for small samples, as expected.

Mean parameter estimates for the regularized models are compared to the population model, which was derived from a standardized 2009 US Sample of PISA ($N = 5,233$). We chose the estimates obtained from this model as the simulation study parameters in order to mimic a real-world model researchers may encounter using these methods so that we can more closely study regularization performance.

The horseshoe priors yield much more shrinkage compared to the ridge and lasso priors, this difference is especially noticeable for small sample sizes. Differences in shrinkage between both versions of the horseshoe prior is minimal.

As with any Bayesian model construction, prior distributions are much less influential when sample sizes are large. Here, we observe this same effect as these regularization priors did not result in much shrinkage with the large sample sizes, even for the horseshoe priors.

Mean LOO-IC estimates are used to compare the predictive performance of the priors of interest in this study. It is important to note that LOO-IC values depend on sample size, as each observation is compared to a predicted outcome and the sum of these differences becomes the LOO estimate. Therefore, comparisons between different sample sizes in predictive performance are impossible using the LOO-IC. Direct comparisons can only be made between models with the same sample size. Results for all four priors and four sample sizes are contained below in Table 4.

These results show that the original horseshoe and the regularized horseshoe prior yield the best predictive performance across sample sizes. The differences between blavaan default non-informative prior and the ridge and lasso priors are smaller, especially when sample sizes are large.

To study the impact of the choice of p_0 on predictive performance, three other p_0 values were used in simulations from the same population path model at the four sample size conditions. Table 5 contains these results. In terms of predictive performance, the selection of p_0 seems to have little impact across varying sample sizes. Next, the shrinkage of coefficients for the regularized horseshoe is shown. The selection of p_0 also did not result in any difference in shrinkage.

These results show that misspecifying p_0 is not detrimental to predictive performance, as differences in LOO-IC estimates between p_0 choices are minimal.

We examined how shrinkage with the regularized horseshoe varied by sample size and prior guess of the number of large coefficients. These results are contained in Figure 8.

For the two smaller sample size conditions, we observe the largest amount of shrinkage when the choice of p_0 is larger however these differences are relatively small. When samples are quite large, the amount of shrinkage across parameters does not vary much as there is much less shrinkage across all model parameters. The choice of p_0 may be slightly more sensitive when the sample size is small, However, this choice does not appear to impact predictive performance.

5.2. Growth Curve Model Simulation Study

5.2.1. Conditions

A Monte Carlo simulation study was conducted to compare the predictive performance and average shrinkage of these regularization priors in the context of Bayesian growth curve modeling. Data were simulated from a similar model as used in the previous growth curve model empirical example, with coefficients artificially magnified by a

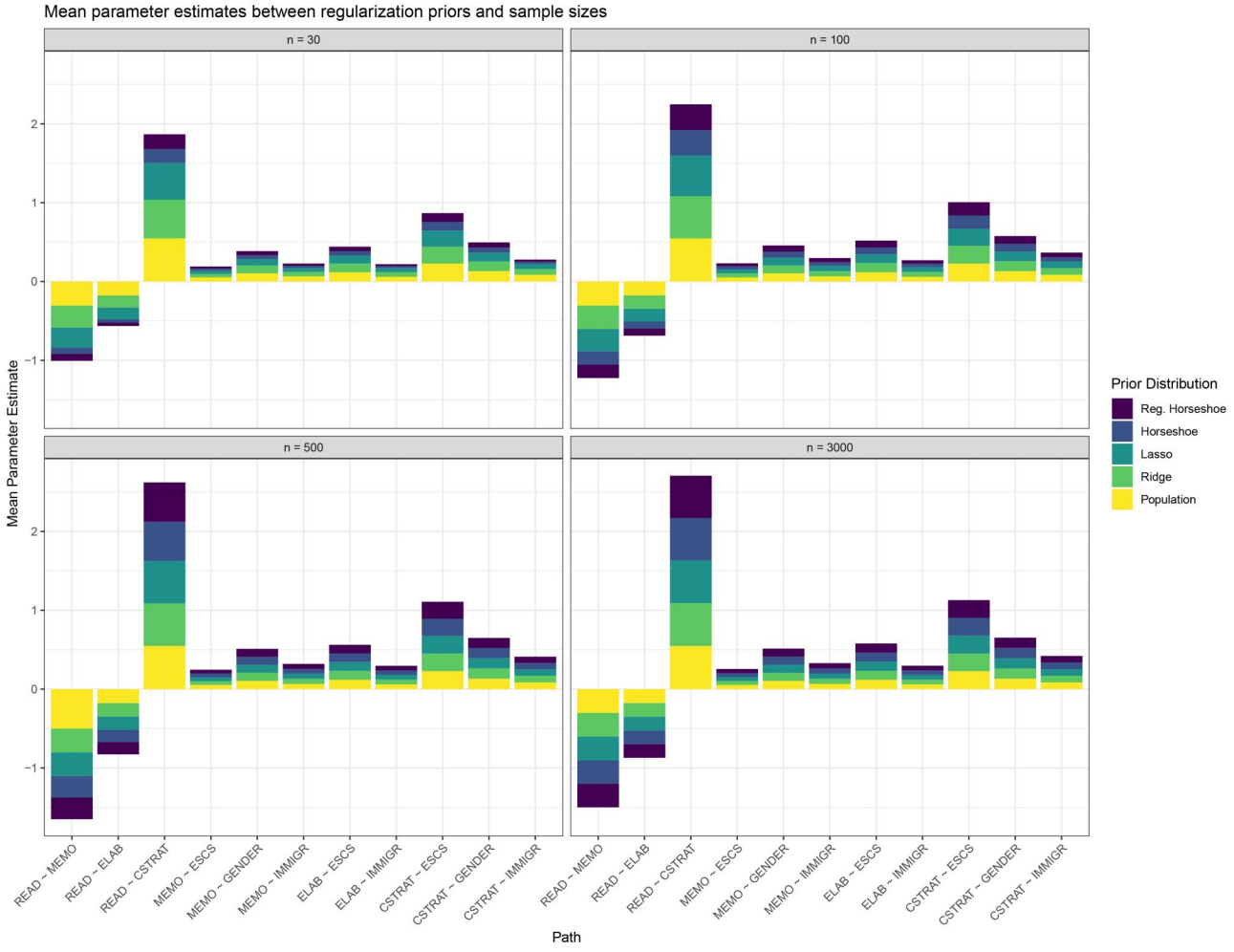

Figure 7. Mean parameter estimates for the given path model in the simulation study for each sample size and prior distribution.

Table 4. Mean LOO-IC estimates for each regularization method.

	n			
	30	100	500	3,000
blavaan default	354.59	1,128.81	5,567.31	33,305.03
Ridge	353.79	1,128.58	5,567.29	33,307.31
Lasso	352.50	1,128.04	5,567.91	33,307.38
Horseshoe	348.75	1,128.67	5,566.76	33,301.90
Regularized horseshoe ($p_0 = 3$)	348.83	1,129.17	5,567.77	33,301.73

Table 5. LOO-IC mean estimates for varying choices of p_0 with the regularized horseshoe.

	n			
p_0	30	100	500	3,000
1	348.73	1,128.94	5,565.76	33,305.34
3	348.83	1,129.17	5,567.77	33,301.73
6	348.70	1,129.16	5,566.02	33,305.36
10	348.65	1,129.23	5,567.04	33,309.99

factor of 10 to better detect shrinkage and model performance. For blavaan defaults and the regularized priors, four different sample size conditions were tested ($n = 50, 100, 500,$ and 1000). Since the previous simulation study, along with previous research has demonstrated the regularized horseshoe's p_0 parameter does not hold a strong influence over the model (Piironen & Vehtari, 2017; van Erp et al., 2019), we will omit testing varying p_0 selections. One thousand replications were generated for each of the 20 conditions of the study. For each replication, models were estimated with Markov chain Monte Carlo sampling using the Gibbs sampler in blavaan with 20,000 iterations with 5,000 adaptation steps. Proper model convergence was reached to allow for further analysis. Software code for the simulation study is available at https://bise.wceruw.org/publications.html.

5.2.2. Results of Simulation Study

Similar to the path analysis study, predictive performance was measured in the form of mean LOO-IC estimates for each simulation condition. We found that the horseshoe priors possess the strongest predictive performance, especially for small samples (Table 6).

Although the differences in LOO-IC estimates may not appear large, any gain in predictive performance (i.e., lower LOO-IC) for a given model is an improvement over competing models. All the regularization priors used in this simulation study slightly out-preformed blavaan default non-informative priors, especially for small samples, we observe the greatest gains in predictive performance with the horseshoe priors. We did not observe much difference between the two horseshoe priors, in fact, the original horseshoe possessed the best LOO-IC estimate for our smallest sample size condition.

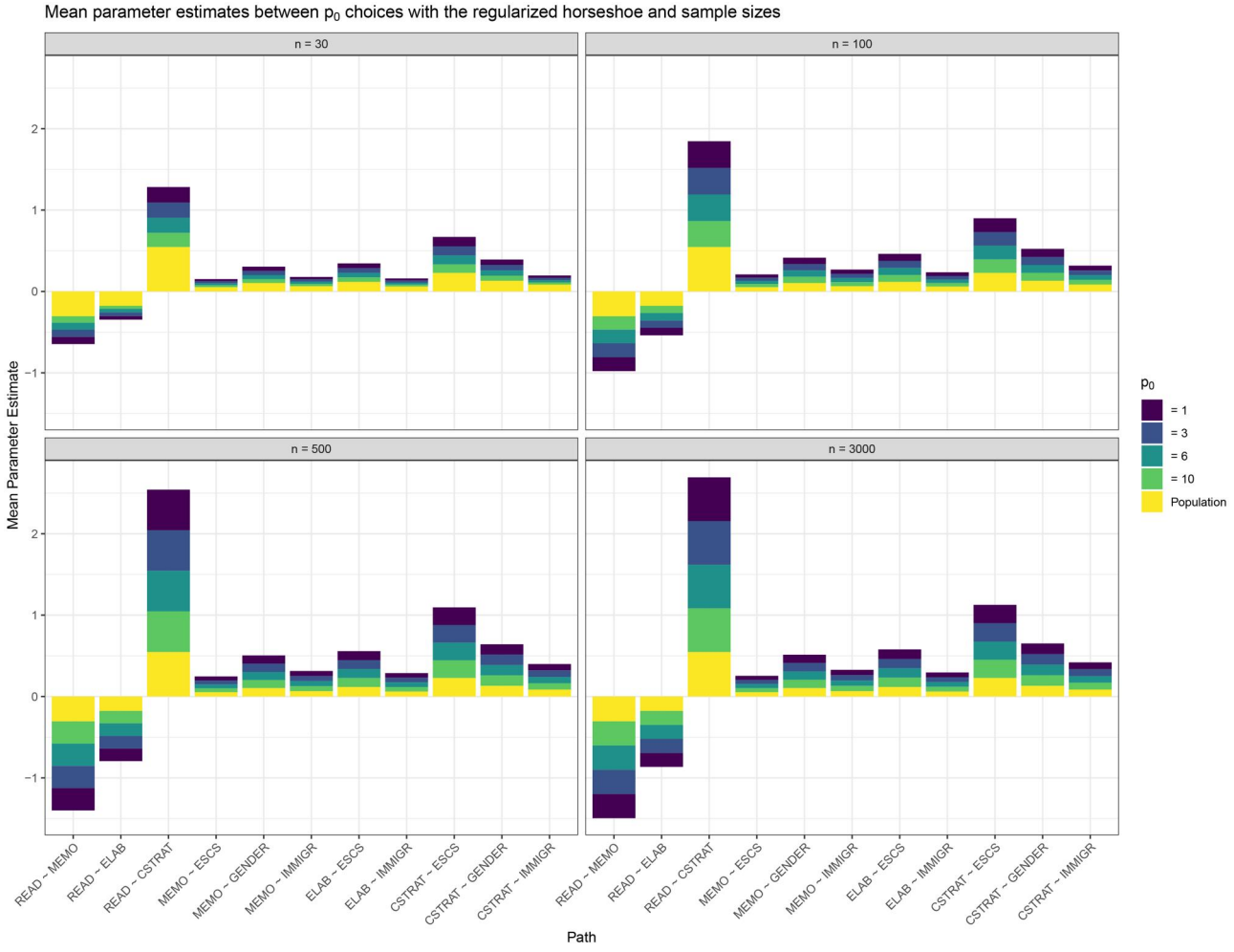

Figure 8. Mean parameter estimates for the given path model in the simulation study for the regularized horseshoe prior with varying choices of p_0 across sample sizes.

Table 6. Mean LOO-IC estimates for each regularization method in the growth curve modeling simulation study.

	n			
	50	100	500	1,000
blavaan default	1,537.96	3,051.71	15,153.97	30,286.84
Ridge	1,537.64	3,051.65	15,153.92	30,287.02
Lasso	1,537.20	3,051.35	15,153.83	30,286.47
Horseshoe	1,535.48	3,049.88	15,153.52	30,286.82
Regularized horseshoe ($p_0 = 1$)	1,535.66	3,050.02	15,153.73	30,286.74

Alongside predictive performance, we also compared shrinkage across regularization priors and sample sizes. Parameter estimates for the population model were derived from NAEP cycles 2003–2017 in order to simulate a realistic model researchers may encounter in their own work. Mean parameter estimates for each regularization prior and sample size are compared to the population model in terms of shrinkage. Figure 9 visually depicts this shrinkage. Similarly to the path analysis simulation, we found that the horseshoe priors induced the greatest amount of shrinkage, particularly when the sample is small. This effect is most notable with the two larger parameters "totrevdiff" and "nslplunchdiff." While the ridge and lasso also shrink parameters closer to zero, shrinkage is more pronounced with the two horseshoe priors for small samples. We did not observe a noticeable difference in terms of shrinkage between the two horseshoe priors.

6. Discussion

The aim of this paper was to assess the performance of the original horseshoe prior and the recently developed regularized horseshoe prior in the context of path analysis and growth curve modeling. These two horseshoe priors were compared to each other and to the ridge and lasso priors in terms of the amount of sparsity induced by the priors and especially in terms of predictive performance as measured by leave-one-out cross-validation. Focusing on the regularized horseshoe, we also examined in path analysis the impact on the variance term, τ_0, as a function of the number of assumed large coefficients in the model, p_0. A further feature of this paper was in demonstrating how to implement these priors in blavaan.

Overall, our results showed that there is little difference between the original horseshoe prior and regularized horseshoe prior in terms of both the amount of sparsity induced in the parameters and predictive performance, but both priors perform noticeably better than the ridge and lasso priors in terms of sparsity and predictive performance. In addition, in the context of these models, the ridge and lasso priors did not perform noticeably better than the case of the non-informative priors in which no sparsity was induced. These findings were true for both the path analysis and the growth

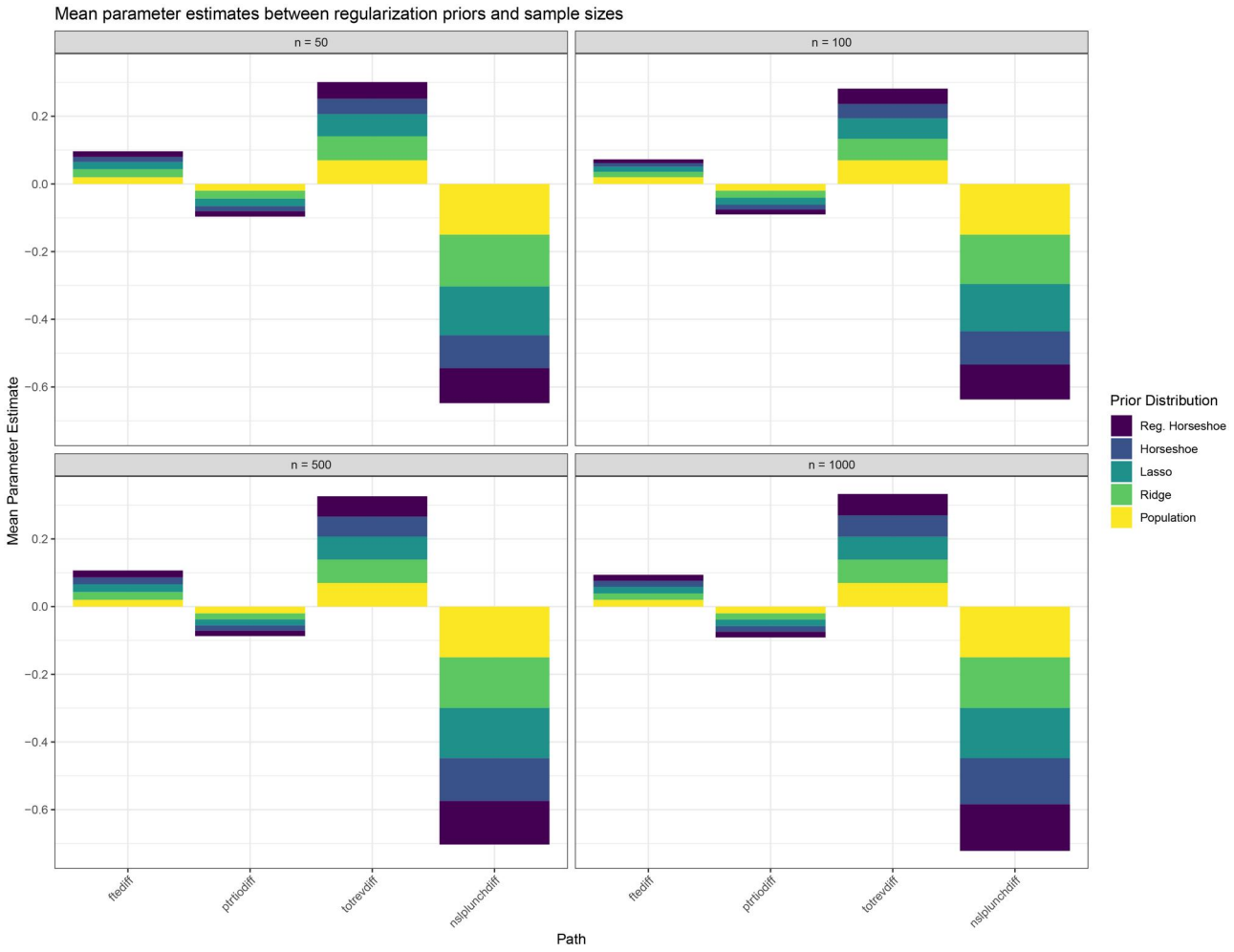

Figure 9. Mean parameter estimates for the given growth curve model in the simulation study. Time-invariant predictors are simulated as the difference between 2017 and 2003 state-level full-time equivalent teachers (ftediff), the difference between 2017 and 2003 pupil-to-teacher ratio (ptrtiodiff), the difference between 2015 and 2003 total revenue, (totrevdiff) and the difference in the 2017 and 2003 percentage of NSLP-eligible students (nslplunchdiff).

curve model studies. Our results are in line with previous findings demonstrating that horseshoe regularization generally provides stronger predictive performance than standard methods of regularization like the ridge or lasso in other contexts such as linear and logistic regression as well as dynamic social network analysis (Carvalho et al., 2009, 2010a; Karimova et al., 2023; Piironen & Vehtari, 2017). Our paper adds to this literature by examining predictive performance in terms of leave-one-out cross-validation—arguably the best scoring rule for true predictive performance.

The present study also assessed the impact of various choices of p_0 on both parameter shrinkage and predictive performance for the regularized horseshoe prior in the context of path analysis. We found the choice of p_0, at least with respect to the model used in this study, to be inconsequential in terms of predictive performance. This finding suggests that the misspecification of this value will not impact predictive accuracy. These results fall in line with previous research on this parameter choice in other statistical contexts such as linear and logistic regression (Piironen & Vehtari, 2017; van Erp et al., 2019).

It should be noted that selecting a single model for the purposes of prediction, even one that has been regularized, is not optimal, and instead many scholars have advocated for some form of model combination. Methods such as Bayesian model averaging (see, e.g., Clyde & George, 2004; Draper, 1995, 2013; Hoeting et al., 1999; Leamer, 1978; Raftery, 1995) and Bayesian stacking (Yao et al., 2018) have been shown to be better suited for the purposes of prediction, but those methods are not the focus of this paper. For an overview of these methods with applications education, see Kaplan (2021), and for a specific extension and application to path analysis, see Kaplan and Lee (2015).

In conclusion, as Bayesian methods are becoming increasingly feasible for researchers to implement in their own analyses, and also be extended to more complex models such as structural equation models (Depaoli, 2021; Depaoli et al., 2023); works such as this paper seek to evaluate the tools available for researchers to use. Our overall conclusion is that the horseshoe priors perform at least as well, if not better than the ridge and lasso priors in terms of predictive performance, and much better in terms of inducing sparsity. However, in any given application it may be reasonable to assess a range of regularization priors in terms of sparsity and predictive performance and choose the one that provides the best performance. Regardless, this paper, and the work of others, suggest that if the goal is to induce

sparsity in structural equation models while achieving good predictive performance, the horseshoe priors are clearly a reasonable choice.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory*. Akademiai Kiado.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, *16*, 125–127. https://doi.org/10.1080/00401706.1974.10489157

Bainter, S. A., McCauley, T. G., Fahmy, M. M., Goodman, Z. T., Kupis, L. B., & Rao, J. S. (2023). Comparing Bayesian variable selection to lasso approaches for applications in psychology. *Psychometrika*, *88*, 1032–1055. https://doi.org/10.1007/s11336-023-09914-9

Betancourt, M. (2018a). *Bayes sparse regression*. https://betanalpha.github.io/assets/case/_studies/bayes/_sparse/_regression.html

Betancourt, M. (2018b). *A conceptual introduction to Hamiltonian Monte Carlo*. https://arxiv.org/pdf/1701.02434.pdf

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley and Sons.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411. https://doi.org/10.32614/RJ-2018-017

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455. https://doi.org/10.2307/1390675

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Proceedings of the twelth international conference on artificial intelligence and statistics* (pp. 73–80). PMLR.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010a). The horseshoe estimator for spare signals. *Biometrika*, *97*, 465–480. https://doi.org/10.1093/biomet/asq017

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010b). The horseshoe estimator for sparse signals. *Biometrika*, *97*, 465–480. https://doi.org/10.1093/biomet/asq017

Clyde, M. A., & George, E. I. (2004). Model uncertainty. *Statistical Science*, *19*, 81–84. https://doi.org/10.1214/088342304000000035

Common Core of Data. (2020). https://nces.ed.gov/ccd/.

Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association*, *77*, 605–610. https://doi.org/10.1080/01621459.1982.10477856

Dawid, A. P. (1984). Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Series A*, *147*, 278–202. https://doi.org/10.2307/2981683

Depaoli, S. (2021). *Bayesian structural equation modeling*. Guilford Publications.

Depaoli, S., Kaplan, D., & Winter, S. (2023). Foundations and extensions of Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed.). Guilford Publications, Inc.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society (Series B)*, *57*, 55–98.

Draper, D. (2013). Bayesian model specification: Heuristics and examples. In *Bayesian theory and applications* (pp. 483–498). Oxford University Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 997–1016. https://doi.org/10.1007/s11222-013-9416-2

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*, 555. https://doi.org/10.3390/e19100555

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889. https://doi.org/10.1080/01621459.1993.10476353

Goodrich, B., Gabry, J., Ali, I., Brilleman, S. (2022). *rstanarm: Bayesian applied regression modeling via Stan* (R package version 2.21.3). https://mc-stan.org/rstanarm/

Gronau, Q. F., & Wagenmakers, E. J. (2019). Limitations of Bayesian leave-one-out cross-validation for model selection. *Computational Brain & Behavior*, *2*, 35–47. https://doi.org/10.1007/s42113-018-0022-4

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoerl, R. W. (1985). Ridge analysis 25 years later. *The American Statistician*, *39*, 186–192. https://doi.org/10.2307/2683926

Hoeting, J. A., Madigan, D., Raftery, A., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382–417. https://doi.org/10.1214/ss/1009212519

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623. http://jmlr.org/papers/v15/hoffman14a.html

Hollenbach, F. M., & Montgomery, J. M. (2020). Bayesian model selection, model comparison, and model averaging. In *The SAGE handbook of research methods in political science and international relations* (pp. 937–960). Sage.

Hsiang, T. C. (1975). A Bayesian view on ridge regression. *The Statistician*, *24*, 267–268. https://doi.org/10.2307/2987923

Jacobucci, R., & Grimm, K. J. (2018). Comparison of frequentist and Bayesian regularization in structural equation modeling. *Structural Equation Modeling*, *25*, 639–649. https://doi.org/10.1080/10705511.2017.1410822

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Sage Publications.

Kaplan, D. (2021). On the quantification of model uncertainty: A Bayesian perspective. *Psychometrika*, *86*, 215–238. https://doi.org/10.1007/s11336-021-09754-5

Kaplan, D. (2023). *Bayesian statistics for the social sciences* (2nd ed.). Guilford Press.

Kaplan, D., & Lee, C. (2015). Bayesian model averaging over directed acyclic graphs with implications for the predictive performance of structural equation models. *Structural Equation Modeling*, *23*, 343–353.

Karimova, D., Leenders, R. T. A., Meijerink-Bosman, M., & Mulder, J. (2023). Separating the wheat from the chaff: Bayesian regularization in dynamic social networks. *Social Networks*, *74*, 139–155. https://doi.org/10.1016/j.socnet.2023.02.006

Leamer, E. E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. Wiley.

Lindley, D. V. (2007). *Understanding uncertainty*. Wiley.

Merkle, E. C., Fitzsimmons, E., Uanhoro, J., & Goodrich, B. (2021). Efficient Bayesian structural equation modeling in Stan. *Journal of Statistical Software*, *100*, 1–22. https://doi.org/10.18637/jss.v100.i06

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, *85*, 1–30. https://doi.org/10.18637/jss.v085.i04

Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*, 1023–1032. https://doi.org/10.1080/01621459.1988.10478694

OECD. (2010). *Pisa 2009 results: Executive summary*. OECD.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*, 681–686. https://doi.org/10.1198/016214508000000337

Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, *11*, 5018–5051. https://doi.org/10.1214/17-EJS1337SI

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. Technische Universität Wien,

Vienna, Austria. https://www.R-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.
Polson, N. G., & Scott, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In *Bayesian statistics 9*. Oxford University Press.
Raftery, A. (1995). Bayesian model selection in social research (with discussion). In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 25, pp. 111–196). Blackwell. https://doi.org/10.2307/271063
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*, 583–639. https://doi.org/10.1111/1467-9868.00353
Stan Development Team. (2021). *Stan modeling language users guide and reference manual, version 2.26* [Computer software manual]. https://mc-stan.org.
Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*, 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x
Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*, 44–47. https://doi.org/10.1111/j.2517-6161.1977.tb01603.x
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2022). National Assessment of Educational Progress NAEP. (https://nces.ed.gov/nationsreportcard/)
van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistical modelling. *Nature Reviews Methods Primers*, *1*, 1–26. https://doi.org/10.1038/s43586-020-00001-2
van Erp, S. (2020). A tutorial on Bayesian penalized regression with shrinkage priors for small sample sizes. In R. van de Schoot & M. Miočević (Eds.), *Small sample size solutions* (pp. 71–84). Taylor and Francis.
van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, *89*, 31–50. https://doi.org/10.1016/j.jmp.2018.12.004
Vehtari, A., Gabry, J., Yao, Y., Gelman, A. (2019). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models* (R package version 2.1.0). https://CRAN.R-project.org/package=loo
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4
Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.
Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*, 917–1007. https://doi.org/10.1214/17-BA1091
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429. https://doi.org/10.1198/016214506000000735
Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix A.

Path Analysis

```
library(lavaan)
library(blavaan)
library(runjags)
library(rjags)

#default BSEM
pathmodel_default <- '
READING ~ MEMO + ELAB + CSTRAT
MEMO ~ ESCS + GENDER + IMMIGR
ELAB ~ ESCS + GENDER + IMMIGR
CSTRAT ~ ESCS + GENDER + IMMIGR
'
fit_default <- bsem(pathmodel_default, data=pisa1, n.chains = 4, sample = 10,000,
adapt = 5,000, burnin = 5,000, target="jags", seed=seed.list)
## BLAVAAN CODE FOR RIDGE PRIOR
pathmodel_ridge <- '
READING ~ prior("dnorm(0,1)")*MEMO +
          prior("dnorm(0,1)")*ELAB +
          prior("dnorm(0,1)")*CSTRAT
MEMO ~ prior("dnorm(0,1)")*ESCS +
          prior("dnorm(0,1)")*GENDER +
          prior("dnorm(0,1)")*IMMIGR
ELAB ~ prior("dnorm(0,1)")*ESCS +
          prior("dnorm(0,1)")*GENDER +
          prior("dnorm(0,1)")*IMMIGR
CSTRAT ~ prior("dnorm(0,1)")*ESCS +
          prior("dnorm(0,1)")*GENDER +
          prior("dnorm(0,1)")*IMMIGR
'
fit_ridge <- bsem(pathmodel_ridge, data=pisa1, n.chains = 4, sample = 10,000,
adapt = 5,000, burnin = 5,000, target ="jags", seed=seed.list)
## BLAVAAN CODE FOR LASSO PRIOR
pathmodel_lasso <- '
READING ~ prior("ddexp(0,1)")*MEMO +
          prior("ddexp(0,1)")*ELAB +
          prior("ddexp(0,1)")*CSTRAT
MEMO ~ prior("ddexp(0,1)")*ESCS +
          prior("ddexp(0,1)")*GENDER +
          prior("ddexp(0,1)")*IMMIGR
ELAB ~ prior("ddexp(0,1)")*ESCS +
          prior("ddexp(0,1)")*GENDER +
          prior("ddexp(0,1)")*IMMIGR
CSTRAT ~ prior("ddexp(0,1)")*ESCS +
          prior("ddexp(0,1)")*GENDER +
          prior("ddexp(0,1)")*IMMIGR
'
fit_lasso <- bsem(pathmodel_lasso, data=pisa1, n.chains = 4, sample = 10,000,
adapt = 5,000, burnin = 5,000, target="jags", seed=seed.list)
## BLAVAAN CODE FOR HORSESHOE PRIOR
pathmodel_horse <- '
READING ~ prior("dnorm(0,tau*lambda_p)")*MEMO +
          prior("dnorm(0,tau*lambda_p)")*ELAB +
          prior("dnorm(0,tau*lambda_p)")*CSTRAT
MEMO ~ prior("dnorm(0,tau*lambda_p)")*ESCS +
          prior("dnorm(0,tau*lambda_p)")*GENDER +
          prior("dnorm(0,tau*lambda_p)")*IMMIGR
ELAB ~ prior("dnorm(0,tau*lambda_p)")*ESCS +
          prior("dnorm(0,tau*lambda_p)")*GENDER +
          prior("dnorm(0,tau*lambda_p)")*IMMIGR
CSTRAT ~ prior("dnorm(0,tau*lambda_p)")*ESCS +
          prior("dnorm(0,tau*lambda_p)")*GENDER +
          prior("dnorm(0,tau*lambda_p)")*IMMIGR
'
extra_horse <-' lambda_p = 1/lambda_p_inv
          lambda_p_inv ~ dt(0,1,1)T(0,) # This induces the half-Cauchy prior.
```

```
        tau = 1/tau_inv
        tau_inv ~ dt(0,1,1)T(0,)
'
fit_horse <- bsem(pathmodel_horse, data=pisa1,
          n.chains = 4, sample = 10,000,
        adapt = 5,000, burnin = 5,000, target =
        "jags",
        mcmcextra=list(syntax=extra_horse,
        monitor=c("lambda_p",          "tau")),
        seed=seed.list)
# BLAVAAN CODE FOR REGUARIZED HORSESHOE PRIOR
(p_0=1)
pathmodel_reghorse1 <- '
READING~prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*MEMO +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*ELAB +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*CSTRAT
MEMO~prior("dnorm(0,tau_inv*lambda_tilde_
      inv)")*ESCS +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*GENDER +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*IMMIGR
ELAB~prior("dnorm(0,tau_inv*lambda_tilde_
       inv)")*ESCS +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*GENDER +
          prior("dnorm(0,tau_inv*lambda_tilde_
          inv)")*IMMIGR
CSTRAT~prior("dnorm(0,tau_inv*lambda_tilde_
       inv)")*ESCS +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*GENDER +
        prior("dnorm(0,tau_inv*lambda_tilde_
        inv)")*IMMIGR
'
extra_reghorse <-' c2 = 1/c2_inv
                   c2_inv ~ dgamma(2, 8)
                   lambda ~ dt(0,1,1)T(0,)
                   tau ~ dt(0,tau0,1)T(0,)
                   tau_inv = 1/tau
                   tau0 = (p0 / (12 - p0)) * (sigma /
                   sqrt(1.0 * n))
                   n = N
                   p0 = 1
                   sigma = 1
                   lambda_tilde_inv = 1/(sqrt(c2)
                   * lambda /
                   sqrt(c2 + tau^2 * lambda^2))
'
fit_reghorse1    <-    bsem(pathmodel_reghorse1,
                 data=pisa1,
                  n.chains = 4, sample = 10,000,
                  adapt = 5,000, burnin = 5,000,
                  target   =   "jags",   mcmcex-
                  tra=list(syntax=extra_
                  reghorse,
                  monitor=c("lambda_tilde_inv",
                  "tau_inv")), seed=seed.list)
```

Growth Curve Modeling

```
library(lavaan)
library(blavaan)
# Default GCM
BayesGCM <- '
# intercept and slope with fixed coefficients
int        =~        1*Math03 + 1*Math05 + 1*Math07
+ 1*Math09 + 1*Math11 +
1*Math13 + 1*Math15 + 1*Math17
slp                                              =~
0*Math03 + 2*Math05 + 4*Math07 + 6*Math09 + 8*Math11 +
10*Math13 + 12*Math15 + 14*Math17
# time-varying covariates
Math03 ~ Reading03
Math05 ~ Reading05
Math07 ~ Reading07
Math09 ~ Reading09
Math11 ~ Reading11
Math13 ~ Reading13
Math15 ~ Reading15
Math17 ~ Reading17
# time-invariant covariates
int ~ prior("dnorm(0,.1)")*1
slp ~ FTEDiff + PTRatioDiff + TOTREVDiff + NSLPLunch-
Diff
'
fitMath <- bgrowth(BayesGCM, data = NAEPdata.std,
target = "jags", n.chains = 4, adapt = 5,000,
sample = 20,000, seed=seed.list, save.lvs=T)
# Ridge
BayesGCMRidge <- '
# intercept and slope with fixed coefficients
int                                              =~
1*Math03 + 1*Math05 + 1*Math07 + 1*Math09 + 1*Math11 +
1*Math13 + 1*Math15 + 1*Math17
slp                                              =~
0*Math03 + 2*Math05 + 4*Math07 + 6*Math09 + 8*Math11 +
10*Math13 + 12*Math15 + 14*Math17
# time-varying covariates
Math03 ~ Reading03
Math05 ~ Reading05
Math07 ~ Reading07
Math09 ~ Reading09
Math11 ~ Reading11
Math13 ~ Reading13
Math15 ~ Reading15
Math17 ~ Reading17
# time-invariant covariates
int ~ prior("dnorm(0,.1)")*1
slp ~ prior("dnorm(0,1)")*FTEDiff +
     prior("dnorm(0,1)")*PTRatioDiff +
     prior("dnorm(0,1)")*TOTREVDiff +
     prior("dnorm(0,1)")*NSLPLunchDiff
'
fitMathRidge   <-   bgrowth(BayesGCMRidge,   data=
NAEPdata.std,
target = "jags", n.chains = 4, adapt = 5,000,
sample = 20,000, seed=seed.list, save.lvs=T)
# Lasso
BayesGCMLasso <- '
# intercept and slope with fixed coefficients
int    =~    1*Math03 + 1*Math05 + 1*Math07 + 1*Math09
        + 1*Math11 +
          1*Math13 + 1*Math15 + 1*Math17
slp    =~    0*Math03 + 2*Math05 + 4*Math07 + 6*Math09
        + 8*Math11 +
       10*Math13 + 12*Math15 + 14*Math17
# time-varying covariates
Math03 ~ Reading03
Math05 ~ Reading05
Math07 ~ Reading07
Math09 ~ Reading09
Math11 ~ Reading11
Math13 ~ Reading13
Math15 ~ Reading15
Math17 ~ Reading17
```

```
# time-invariant covariates
int ~ prior("dnorm(0,.1)")*1
slp ~ prior("ddexp(0,1)")*FTEDiff +
    prior("ddexp(0,1)")*PTRatioDiff +
    prior("ddexp(0,1)")*TOTREVDiff +
    prior("ddexp(0,1)")*NSLPLunchDiff
'
fitMathLasso <- bgrowth(BayesGCMLasso,
data = NAEPdata.std,
         target = "jags", n.chains = 4, adapt =
         5,000,
         sample = 20,000, seed = seed.list,
         save.lvs = T)
# Horseshoe
BayesGCMHorseshoe <- '
# intercept and slope with fixed coefficients
int  =~  1*Math03 + 1*Math05 + 1*Math07 + 1*Math09
       + 1*Math11 +
         1*Math13 + 1*Math15 + 1*Math17
slp  =~  0*Math03 + 2*Math05 + 4*Math07 + 6*Math09
       + 8*Math11 +
         10*Math13 + 12*Math15 + 14*Math17
# time-varying covariates
Math03 ~ Reading03
Math05 ~ Reading05
Math07 ~ Reading07
Math09 ~ Reading09
Math11 ~ Reading11
Math13 ~ Reading13
Math15 ~ Reading15
Math17 ~ Reading17
# time-invariant covariates
int ~ prior("dnorm(0,.1)")*1
slp ~ prior("dnorm(0,tau*lambda_p)")*FTEDiff +
    prior("dnorm(0,tau*lambda_p)")*PTRatioDiff +
    prior("dnorm(0,tau*lambda_p)")*TOTREVDiff +
    prior("dnorm(0,tau*lambda_p)")*NSLPLunchDiff
'
extra_horse <-' lambda_p = 1/lambda_p_inv
         lambda_p_inv ~ dt(0,1,1)T(0,)
         tau = 1/tau_inv
         tau_inv ~ dt(0,1,1)T(0,)
'
fitMathHorseshoe <- bgrowth(BayesGCMHorseshoe,
data = NAEPdata.std,
target = "jags", mcmcextra = list(syntax = extra_
horse,
monitor = c("lambda_p", "tau")),
n.chains = 4, adapt = 5,000,
sample = 20,000, seed = seed.list, save.lvs = T)
# Regularized Horseshoe
BayesGCMRegHorse <- '
# intercept and slope with fixed coefficients
int  =~  1*Math03 + 1*Math05 + 1*Math07 + 1*Math09
       + 1*Math11 +
       1*Math13 + 1*Math15 + 1*Math17
slp  =~  0*Math03 + 2*Math05 + 4*Math07 + 6*Math09
       + 8*Math11 +
       10*Math13 + 12*Math15 + 14*Math17
# time-varying covariates
Math03 ~ Reading03
Math05 ~ Reading05
Math07 ~ Reading07
Math09 ~ Reading09
Math11 ~ Reading11
Math13 ~ Reading13
Math15 ~ Reading15
Math17 ~ Reading17
# time-invariant covariates
int ~ prior("dnorm(0,.1)")*1
slp ~ prior("dnorm(0,tau_inv*lambda_tilde_
inv)")*FTEDiff +
prior("dnorm(0,tau_inv*lambda_tilde_
inv)")*PTRatioDiff +
prior("dnorm(0,tau_inv*lambda_tilde_
inv)")*TOTREVDiff +
prior("dnorm(0,tau_inv*lambda_tilde_
inv)")*NSLPLunchDiff
'
N <- nrow(NAEPdata.std)
extra_reghorse <- '
c2 = 1/c2_inv
c2_inv ~ dgamma(2, 8)
lambdatest ~ dt(0, 1, 1) T(0,)
tau ~ dt(0, tau0, 1) T(0,)
tau_inv = 1/tau
tau0 = (p0 / (p - p0)) * (sigma / sqrt(1.0 * n))
n = N
p0 = 1
p = 5
sigma = 1
lambda_tilde_inv = 1 / (sqrt(c2) * lambdatest /
sqrt(c2 + tau^2 * lambdatest^2))
'
fitMathRegHorse <- bgrowth(BayesGCMRegHorse,
                  data = NAEPdata.std,
                  mcmcextra = list(syntax = extra_
                  reghorse,
                  monitor = c("lambda_tilde_inv",
                  "tau_inv")),
                  target = "jags", n.chains = 4,
                  adapt = 5,000,
                  sample = 20,000, seed = seed.-
                  list, save.lvs = T)
```