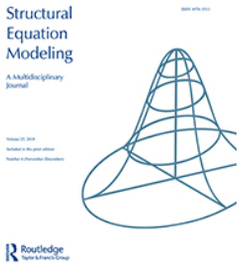
Routledge
Taylor & Francis Group



Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

An Adaptive Bayesian Lasso Approach with Spike-and-Slab Priors to Identify Multiple Linear and Nonlinear Effects in Structural Equation Models

Holger Brandt, Jenna Cambria & Augustin Kelava

To cite this article: Holger Brandt, Jenna Cambria & Augustin Kelava (2018) An Adaptive Bayesian Lasso Approach with Spike-and-Slab Priors to Identify Multiple Linear and Nonlinear Effects in Structural Equation Models, Structural Equation Modeling: A Multidisciplinary Journal, 25:6, 946-960, DOI: 10.1080/10705511.2018.1474114

To link to this article: https://doi.org/10.1080/10705511.2018.1474114

Structural Equation Modeling: A Multidisciplinary Journal, 25: 946–960, 2018

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1474114

An Adaptive Bayesian Lasso Approach with Spike-and-Slab Priors to Identify Multiple Linear and Nonlinear Effects in Structural Equation Models

Holger Brandt,[1] Jenna Cambria,[2] and Augustin Kelava[3]

[1]*University of Kansas*
[2]*University of Arkansas*
[3]*Eberhard Karls University Tuebingen*

In applied research, such as with motivation theories, typically many variables are theoretically implied predictors of an outcome and several interactions are assumed (e.g., Watt, 2004). However, estimation problems that might arise when several interaction and/or quadratic effects are analyzed simultaneously have not been investigated because simulation studies on interaction effects in the structural equation modeling framework have mainly focused on small models that contain single interaction effects. In this article, we show that traditional approaches can provide estimates with low accuracy when complex models are estimated. We introduce an adaptive Bayesian lasso approach with spike-and-slab priors that overcomes this problem. Using a complex model in a simulation study, we show that the parameter estimates of the proposed approach are more accurate in situations with high multicollinearity or low reliability compared with a standard Bayesian lasso approach and typical frequentist approaches (i.e., unconstrained product indicator approach and latent moderated structures approach).

Keywords: Adaptive lasso, continuous mixture, interaction effect, multicollinearity, quadratic effect, regularization, reliability, spike-and-slab priors

Many theories in applied research propose complex patterns of differential relationships, such as expectancy values theory (Eccles et al., 1983), theories on the prediction of educational choices using vocational interests (Lapan & Jingelski, 1992), or theories on burnout (e.g., Hao, Hong, Xu, Zhou, & Xie, 2015). For example, in expectancy-value theory, which is a dominant theory in educational psychology, expectancies for success and subjective task values are seen as energizers of action and as the direct predictors of achievement and choice outcomes (Wigfield & Cambria, 2010a). Expectancies for success and values include four major components, namely attainment value, intrinsic value, utility value, and cost (Wigfield & Cambria, 2010b), which are influenced by a series of psychological, social, contextual, and sociocultural influences (e.g., Wigfield & Eccles, 2000). Not only are expectancies and values significantly mutually correlated (e.g., Watt, 2004), but there has also been a renewed focus on examining the interaction between expectancies and values in predicting engagement, career intentions, and achievement (e.g., Trautwein et al., 2012).

In many empirical research articles using nonlinear structural equation models (SEM), these theories are typically examined using smaller models that include only single interaction effects (e.g., Dakanalis et al., 2014; Dicke et al., 2014; Scalco et al., 2014). Instead of a complex model that accounts for all relevant coexisting constructs and their nonlinear effects simultaneously, often several analyses are conducted separately for each of the interaction effects under investigation (e.g., J. Guo, Marsh, Parker, Morin, & Yeung, 2015; Trautwein et al., 2012). These models might be preferred because complex models tend to suffer from problems such as inaccurate

Correspondence should be addressed to Holger Brandt, Department of Psychology, University of Kansas, 1450 Jayhawk Blvd, Lawrence, KS 66045. E-mail: holger.brandt@ku.edu

estimates or low convergence rates. However, they may not reflect reality or the complexity of the theory, which may lead to statistical artifacts.

ESTIMATORS FOR NONLINEAR SEM

Several different statistical approaches have been developed that account for multiple interaction or quadratic effects in nonlinear SEM (e.g., Jöreskog & Yang, 1996; Kelava & Brandt, 2009; Kelava, Nagengast, & Brandt, 2014; Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Marsh, Wen, & Hau, 2004; Schumacker & Marcoulides, 1998; Wall & Amemiya, 2003). In general, latent interaction effects are used to investigate differential hypotheses; that is, they test whether relationships between predictors and outcomes depend on a third variable. Quadratic effects are used to model curvilinear relationships. Several researchers have advocated for them to be included routinely in interaction models in order to reduce statistical artifacts (e.g., Ganzach, 1997; Klein, Schermelleh-Engel, Moosbrugger, & Kelava, 2009). However, most simulation studies that use these approaches have primarily been conducted using basic models that only include single interaction effects and sometimes quadratic effects (e.g., Brandt, Kelava, & Klein, 2014; Cham, West, Ma, & Aiken, 2012; Kelava et al., 2011; Marsh et al., 2004; and see Supplement 1[1] for a list of recent simulation studies). As a consequence, it remains unclear how these approaches perform when models become more complex; that is, when more than two or three interaction effects and/or quadratic effects are investigated simultaneously. In these situations, we can expect that the problem of multicollinearity becomes more prominent because the product terms used to operationalize interaction or quadratic effects can also be highly correlated.

Alternative approaches, such as the lasso approach (Tibshirani, 1996), have been developed in a regression framework for complex models with many predictors and situations with high multicollinearity. Their applicability to nonlinear SEM has not yet been investigated in detail. Here, we propose an adaptive Bayesian lasso approach with spike-and-slab priors (aBSS-lasso) that may have beneficial estimation aspects in complex settings.

ILLUSTRATION OF THE PROBLEM

In order to illustrate that the problem of model complexity is severe, we conducted a small simulation study using the TIMSS 2015 data set (Mullis & Martin, 2013) and applied the expectancy-value theory to predict math achievement (details on the model and data set can be found in Supplement 1, see Footnote 1). We investigated the performance of two standard approaches used for estimating latent interactions, the unconstrained product indicator (UPI) approach (Kelava & Brandt, 2009; Marsh et al., 2004) and the latent moderated structures (LMS) approach (Klein & Moosbrugger, 2000). We compared their performance to the aBSS-lasso proposed in this article. Increasing model complexity was induced by including two, three, or four latent predictors, and their respective interactions and quadratic effects. The first model only included expectancies and intrinsic value (one interaction, two quadratic effects), the second model added utility value (three interactions and three quadratic effects), and the last model added attainment value (six interactions and four quadratic effects). The last model reflected the complexity of expectancy-value theory that makes simultaneous predictions based on all of these coexisting constructs.[2]

We used the US subsample of the data set ($N = 10,163$) and randomly drew 100 samples, each with $N = 400$ students and analyzed them. Similar average parameter estimates were obtained across the models and estimators. However, the variability of the parameter estimates across the sample increased dramatically from the simple two-predictor model to the complex four-predictor model. This increase was strongest for the UPI approach in which the median SD of the parameter estimates for interaction and quadratic effects increased from 0.23 to 0.86 (with a maximum SD of 2.12). For LMS, the median SD increased from 0.19 to 0.25 (with a maximum SD of 0.50). The proposed lasso method provided far more accurate results with a median SD of 0.09 (and a maximum SD of 0.12) for the complex model, which was 2.8 times smaller than for LMS and 9.6 times smaller than for the UPI approach. This implied that two researchers who analyzed two different samples with LMS or the UPI approach might end up with very different results (e.g., with a strong positive or negative or zero interaction effect), whereas with the aBSS-lasso they would have obtained very similar results.

This example illustrates the problem that we want to address in this article: If more complex models are analyzed that include several nonlinear effects that are meaningful from a modeling perspective, the multicollinearity due to the correlations between the latent product terms brings about a loss of estimation accuracy in standard approaches. This problem can be addressed with shrinkage methods, such as the lasso.

[1] The supplement can be downloaded from the first author's personal website www.holger-brandts-methods.com.

[2] False constraints in the structural model (i.e., omitting relevant nonlinear effects) can result in spurious estimates for the remaining parameters and may not show up in model fit (Gerhard, Büchner, Klein, & Schermelleh-Engel, 2014; Gerhard et al., 2014; Klein et al., 2009). From this perspective, it is beneficial to include these nonlinear terms in this model.

Scope of the Article

In this article, we pursue two goals. First, we will propose an aBSS-lasso for nonlinear SEM that has advantages in variable selection and performance compared to other traditional approaches to nonlinear SEM. Second, we will compare the performance of the proposed aBSS-lasso approach to traditional approaches to nonlinear SEM (LMS and UPI) and a standard adaptive Bayesian lasso implementation (aB-lasso) in situations with many predictor variables and increasing multicollinearity. The advantages of the aBSS-lasso are expected to be more prominent with increasing multicollinearity.

Structure of the Article

In the following section, we will introduce the lasso estimator that has been developed to analyze complex models with many predictor variables in regression analysis. We will then extend this model to nonlinear SEM and provide a Bayesian specification, the aBSS-lasso. In a simulation study, we will investigate the performance of the proposed method. Finally, we will discuss the simulation results and give recommendations on the use of the aBSS-lasso.

LASSO

The lasso comprises several different shrinkage methods that were developed in the regression framework. We will review the original lasso and its adaptive version. Then, we provide information on their Bayesian implementations using Laplace priors and its extension to spike-and-slab priors that provide better shrinkage characteristics than the Laplace priors. We discuss the selection of shrinkage parameters and shortly summarize lasso implementations in the latent variable framework.

The lasso estimator was introduced by Tibshirani (1996) as an alternative to the OLS estimator for multiple regression models. This estimator shrinks small regression effects to zero by imposing a penalty term (regularization) to the OLS criterion. For a regression with $u = 1 \ldots m$ standardized predictor variables $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$ and a dependent variable $\mathbf{y}$, the lasso estimates $\hat{\beta} = (\hat{\beta}_1, \ldots \hat{\beta}_m)'$ are given by

$$\hat{\beta} = \min_{\beta} \left\{ (\tilde{\mathbf{y}} - \mathbf{X}\beta)'(\tilde{\mathbf{y}} - \mathbf{X}\beta) + \lambda \sum_{u=1}^{m} |\beta_u| \right\} \qquad (1)$$

with $\tilde{\mathbf{y}} = \mathbf{y} - \beta_0$ (Park & Casella, 2008; Tibshirani, 1996). The lasso estimates can be viewed as L_1-penalized least-squares estimates (Park & Casella, 2008). The shrinkage parameter $\lambda \geq 0$ controls the amount of shrinkage. With $\lambda = 0$, the estimator is equivalent to the OLS estimator. With increasing $\lambda \to \infty$, the amount of shrinkage imposed on the regression coefficients increases until all coefficients are shrunken to zero (different methods on how to select or estimate λ are discussed below).

The advantage of the lasso estimator compared to the OLS estimator is that it is designed to provide smaller standard errors ("super-efficient" estimator; Tibshirani, 1996). Further, the lasso can be used as a selection criterion for predictor variables. Due to the imposed shrinkage, parameters close to zero are shrunken exactly to zero. In situations with many predictor variables, this implies two advantages: first, model interpretation and sparsity are facilitated because the model provides a very clear structure of a subset of variables with no influence at all and a subset of variables with regression coefficients unequal to zero. Second, the problem of multicollinearity is attenuated because the estimates and (semi)-partial correlations are more stable (Zou, 2006).

Besides the L_1 penalty, different versions of penalties have been introduced, such as the ridge regression that uses an L_2 penalty of the form $\lambda \sum_{j=1}^{p} (\beta_j)^2$ (Hoerl & Kennard, 1970), or the generalized L_q penalty called bridge regression (e.g., Frank & Friedman, 1993). Further extensions of the lasso include the adaptive lasso (Griffin & Brown, 2011; Wang, Li, & Tsai, 2007; Zou, 2006), the elastic net (a combination of lasso and ridge regression; Zou & Hastie, 2005), group lasso estimators (a variant particularly for categorical predictor variables; e.g., Feng, Wang, Wang, & Song, 2015; Meier, Van De Geer, & Bühlmann, 2008; Xu & Ghosh, 2015; Yuan & Lin, 2006), and Bayesian implementations of the lasso (e.g., Hans, 2009; Park & Casella, 2008) or the adaptive lasso (Leng, Tran, & Nott, 2014).

Adaptive Lasso

The adaptive lasso allows for a more flexible control of the shrinkage imposed on the parameters by selecting $u = 1 \ldots m$ different shrinkage coefficients (λ_u) for each of the m regression coefficients. The regression coefficients are received from the function

$$\hat{\beta} = \min_{\beta} \left\{ (\tilde{\mathbf{y}} - \mathbf{X}\beta)'(\tilde{\mathbf{y}} - \mathbf{X}\beta) + \sum_{u=1}^{m} \lambda_u |\beta_u| \right\}. \qquad (2)$$

The shrinkage of variables can be controlled, for example, by using smaller shrinkage parameters for important variables and larger ones for unimportant variables (Leng et al., 2014; Zou, 2006).

The adaptive lasso overcomes one of the main problems of the original lasso; that is, the original lasso produces inconsistent parameter estimates in some situations independent of the actual choice of the shrinkage

factor because it shrinks all coefficients with the same shrinkage factor (Zou, 2006). The adaptive lasso can produce consistent estimates with an appropriate choice of shrinkage factors (Leng et al., 2014). Another advantage of the adaptive lasso is that the original limitation of the lasso – that predictor variables should be standardized to ensure a similar penalty on all effects – is attenuated. This is particularly important for situations with models that include interaction effects because products of standardized variables are not standardized themselves. Although product variables in manifest regression analyses can be standardized a priori, this is more complicated in latent variable models where a model implied standardization of latent variables is more complicated (cf. Brandt, Umbach, & Kelava, 2015; Song, Li, Cai, & Ip, 2013).

Bayesian (Adaptive) Lasso

The penalty imposed on the OLS criterion by Equation 1 can also be implemented in a Bayesian framework (e.g., Hans, 2009; Kyung, Gill, Ghosh, & Casella, 2010; Park & Casella, 2008). In a Bayesian lasso regression, each of the regression coefficients follows a conditional double exponential prior (Laplace prior; Park & Casella, 2008):

$$f(\beta|\sigma^2) = \prod_{u=1}^{m} \frac{\lambda}{2\sqrt{\sigma^2}} \exp\left(\frac{-\lambda|\beta_u|}{\sqrt{\sigma^2}}\right) \tag{3}$$

with a marginal prior for the residual variance σ^2 of $\mathbf{y}$. The lasso estimate is equivalent to the mode of the posterior distribution (Tibshirani, 1996).

The adaptive version of the Bayesian lasso uses separate shrinkage factors λ_u for each regression coefficients (Griffin & Brown, 2011; Leng et al., 2014):

$$f(\beta|\sigma^2) = \prod_{u=1}^{m} \frac{\lambda_u}{2\sqrt{\sigma^2}} \exp\left(\frac{-\lambda_u|\beta_u|}{\sqrt{\sigma^2}}\right). \tag{4}$$

Again, it can be expected that the λs for small regression coefficients are larger than those of the larger coefficients (see simulation results in Leng et al., 2014).

One aspect of the Bayesian implementation should be noted: in order to preserve the variable selection property, the mode of the posterior needs to be used. Different methods on how to obtain this mode have been discussed (e.g., Griffin & Brown, 2011; Leng et al., 2014; Yuan & Lin, 2005). Often, the mode is not the preferred statistic in Bayesian modeling either because this value can be unreliable in certain situations (e.g., when the posterior distribution is not unimodal) or because the posterior can provide more information via its mean and credibility intervals (Gelman et al., 2013). If the mean of the posterior is used as an estimate, the Bayesian lasso does not produce exact zeros and loses its variable selection property (Leng et al., 2014). One method that overcomes this disadvantage is the spike-and-slab prior.

Spike-and-Slab Priors

An important extension has been proposed to capture the advantage of the lasso in the Bayesian framework when the number of possible predictors is large: the spike-and-slab prior (e.g., Mitchell & Beauchamp, 1988; Fahrmeir, Kneib, & Konrath, 2010; Yuan & Lin, 2005) and the utilization of a selection variable that controls the inclusion of variables (Dellaportas, Forster, & Ntzoufras, 2002; George & Culloch, 1993; Kuo & Mallick, 1998).

The basic idea of this Bayesian variable selection technique is to combine a prior distribution (e.g., Bernoulli) that controls the inclusion of variables in the regression with a prior distribution (e.g., normal) of those regression coefficients that have effects different from zero (conditional on the subset of the selected variables).

Different types of these models have been proposed that can be grouped, for example, by the type of selection variable. Kuo and Mallick (1998) suggest to use a binary variable ν that controls the inclusion of each predictor variable. They formulate the following model for each of the $u = 1, \ldots, m$ regression coefficients (cf. Lykou & Ntzoufras, 2013):

$$\beta_u = \beta_u^* \nu_u \tag{5}$$

$$\beta_u^* \sim dexp(0, 1/(\sigma\lambda_u)) \tag{6}$$

$$\nu_u \sim Bern(\pi_u) \tag{7}$$

$$\pi_u \sim Beta(a_u, b_u), \tag{8}$$

where $dexp(\mu, \sigma)$ is the double-exponential (Laplace) prior with mean μ and scale σ, and σ^2 is the residual variance of $\mathbf{y}$. ν_u is a binary Bernoulli distributed variable with *Beta* distributed probability π_u that indicates the probability that the regression coefficient is different from zero. This model is basically a discrete mixture model with two components ($\beta_u^* \nu_u + 0(1 - \nu_u)$). The first component follows a double exponential distribution (slab) and the second component is a constant mass at zero (spike). The advantage of the model is that the prior specification leads to distinct subsets of zero and nonzero regression coefficients. Small regression coefficients are shrunken to zero. The disadvantage of this model is that it produces a high-dimensional discrete parameter space, which results in strong computational burden (Bhadra, Datta, Polson, & Willard, 2017).

Ishwaran and Rao (2005) proposed a different model within the ridge regression framework. Instead of sampling the binary variable ν, they formulate a more direct approach:

$$\beta_u \sim N(0, \sqrt{\sigma_1^2 \pi_u + \sigma_2^2 (1 - \pi_u)}) \tag{9}$$

$$\pi_u \sim Beta(a_u, b_u), \tag{10}$$

where $N(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ. This model essentially represents a continuous mixture model for the regression coefficients:

$$\beta_u = \beta_{1u}^* \pi_u + \beta_{2u}^* (1 - \pi_u) \tag{11}$$

$$\beta_{ju}^* \sim N(0, \sigma_j), j = 1, 2 \tag{12}$$

$$\pi_u \sim Beta(a_u, b_u). \tag{13}$$

In this model, the variance σ_2^2 is selected using a very small value, which – similar to the specification above – results in a very spiky distribution concentrated around zero. The variable π_u can be viewed as a complexity parameter (Ishwaran & Rao, 2005) that controls the complexity of the model. If effects are small, π_u approximates zero and hence, coefficients are close to zero, too. If effects are large, π_u approximates one and the size of the effects are primarily driven by β_{1u}^*. This model allows for more flexibility and a higher amount of adaptivity compared to the model in Equation 8 (Ishwaran & Rao, 2005).

Selection of Shrinkage Parameter(s)

For the selection of the shrinkage parameter λ, different procedures have been suggested. These include graphical inspection of the amount of shrinkage for each parameter across different selected shrinkage parameters (regularization plots; Lykou & Ntzoufras, 2013), the evaluation of empirical Bayes approaches (Atchade, 2011; Casella, 2001; Lykou & Ntzoufras, 2013), and cross-validation techniques (e.g., Tibshirani, 1996).

Within the Bayesian framework, the usage of prior distributions for the shrinkage factor λ allows for a more flexible estimation (Guo, Zhu, Chow, & Ibrahim, 2012; Lykou & Ntzoufras, 2013; Park & Casella, 2008). The selection of fixed shrinkage factors, which always implies some subjective judgment by the researcher, is not necessary. In contrast, using uninformative prior distributions with heavy tails for the shrinkage factors allows for a more model-driven selection of the shrinkage (cf. Feng et al., 2015; Feng, Wu, & Song, 2015). Typical distributions used are the inverse gamma distribution or the Half Cauchy distribution (Gelman, 2006; Polson & Scott, 2011). These priors have asymmetrical distributions that are bounded at zero; selecting hyperparameters that allow for large variances of the priors gives the necessary flexibility in the sampling.

Latent Variable Lasso Implementations

Recently, the lasso estimator has been introduced to latent variable models. Feng, Wu, and Song (2015) applied the Bayesian adaptive lasso to latent ordinal regression, extending previous work by Leng et al. (2014). Jacobucci, Grimm, and McArdle (2016) provided an R package called regsem for penalized SEM by applying a penalty directly to the likelihood function:

$$F_{regsem} = F_{ML} + \lambda P(\cdot). \tag{14}$$

This package can be used for linear SEM. Here, $P(\cdot)$ includes the penalty function for the (sum of the) parameters (see Equation 1). Details are provided in Jacobucci et al. (2016). Further extensions include methods for semiparametric SEM that use the (adaptive) Bayesian lasso for latent spline models (Guo et al., 2012), and group lasso implementations (Feng et al., 2015). Implementations in SEM using a spike-and-slab prior have not yet been proposed (or tested in a simulation study).

Conclusions

A majority of developments in the area of penalized models—particularly the lasso can be found for manifest variable models. Extensions to latent models are rare and implementations are seldom available for applied users. In the following section, we will specify a general latent (nonlinear) adaptive Bayesian lasso model with spike-and-slab priors. This general model can be used for linear or for nonlinear SEM; its advantage compared to traditional approaches is more pronounced if many latent variables are used. In a simulation study, we will compare its estimation properties to traditional maximum likelihood approaches that have been used predominantly for nonlinear SEM in the recent past and to a standard adaptive Bayesian lasso.

ADAPTIVE BAYESIAN LASSO WITH SPIKE-AND-SLAB PRIORS FOR SEM: ABSS-LASSO

In this section, we provide the general specification of the aBSS-lasso for SEM. The model is general as it covers linear and nonlinear models. We compare the model specification to a standard adaptive Bayesian lasso implementation (aB-lasso) and give guidelines on the choice of the spike-and-slab priors.

For each subject $i = 1 \ldots N$ the measurement model the p dimensional vector $\mathbf{x}_i$ and the q dimensional vector $\mathbf{y}_i$ is given by

$$\mathbf{x}_i = \tau_x + \mathbf{\Lambda}_x \xi_i + \delta_i \tag{15}$$

$$\mathbf{y}_i = \tau_y + \Lambda_y \eta_i + \in_i, \tag{16}$$

where τ_x and τ_y are p- and q dimensional vectors of intercepts, Λ_x and Λ_y are $p \times m$ and $q \times n$ dimensional factor loading matrices, ξ_i includes the m independent latent variables, η_i includes the n dependent variables, and, finally δ_i and ϵ_i are residual vectors whose elements are assumed to be mutually independent and independent with ξ_i and η_i.

The structural model is given by

$$\eta_i = \alpha + \Gamma_1 \xi_i + \Gamma_2 vech(\xi_i \xi_i') + \zeta_i, \tag{17}$$

where α is a vector of latent intercepts, Γ_1 and Γ_2 include the regression coefficients, and ζ_i is a residual vector. $vech(\xi_i \xi_i')$ is an $m(m+1)/2$ dimensional vector that includes all nonredundant product terms of the latent predictor vector ξ_i.

The structural model as specified in Equation 17 is a saturated model according to the fact that it includes all possible two-way interaction and quadratic effects that can be formulated with the predictor vector. The model can be simplified to include less than $m(m+1)/2$ effects. Or, the model can be reduced to a purely linear model by omitting $\Gamma_2 vech(\xi_i \xi_i')$. The model formulation can be extended to include directed linear and nonlinear effects between the dependent variables. Here, we do without this model part in order to keep a simple model.

Bayesian Model Specification

Except for the spike-and-slab model part that refers to the regression coefficients in the model, the Bayesian model specification follows a standard implementation for SEM and is provided in Supplement 2 (see Footnote 1; see e.g., R. Guo et al., 2012; Song et al., 2013). The spike-and-slab prior for the lasso part of the model is specified using a product of a double exponential (Laplace) distribution and a Beta distributed prior. For each regression coefficient $\gamma_{w,vu}$ $(w = 1, 2)$ in the matrices Γ_1 and Γ_2, the following specification is used for an adaptive lasso implementation

$$\gamma_{w,vu} = \gamma^*_{w,vu} \pi_{w,vu} \tag{18}$$

$$\gamma^*_{w,vu} \sim dexp\big(0, 1/(\sigma_{\zeta vv} \lambda_{w,vu})\big) \tag{19}$$

$$\sigma_{\zeta vv} \sim C^+(\alpha, \beta) \tag{20}$$

$$\lambda^2_{w,vu} \sim C^+(\alpha, \beta) \tag{21}$$

$$\pi_{w,vu} \sim Beta(a_{w,vu}, b_{w,vu}) \tag{22}$$

where $C^+(\alpha, \beta)$ is the Half Cauchy distribution with hyperpriors (α, β), and $Beta(a_{w,vu}, b_{w,vu})$ is the Beta distribution with hyperpriors $a_{w,vu}, b_{w,vu}$.

Properties of the Implementation

For each coefficient, two relevant parameters are specified: first, the shrinkage parameter λ that controls the scale of the double exponential (Laplace) distribution from which the coefficient γ^* is sampled. With increasing λ the variance decreases and approximates zero. This results in small, strongly shrunken effects γ^*. In order to provide flexibility for the sampling of the shrinkage factor, a Half Cauchy distribution is proposed (Park & Casella, 2008; Polson & Scott, 2011).[3] The usage of $\sigma_{\zeta vv}$ in the definition of the prior variance of γ^* allows us to control the shrinkage across different dependent variables with different (residual) variances. When using only the shrinkage due to the double exponential (Laplace) distribution, a standard adaptive Bayesian lasso (aB-lasso) procedure results (i.e., $\gamma_{w,vu} = \gamma^*_{w,vu}$ instead of Equation 18; cf. Feng, Wu, & Song, 2017; Leng et al., 2014). This procedure results in less shrinkage than the spike-and-slab prior.

Second, a continuous variable π is generated from a Beta distribution. This distribution is limited between zero and one. The multiplication of π and γ^* can be viewed as a continuous mixture distribution. Here, we model a mixture of a double exponential (Laplace) distribution and a constant zero mass distribution. If π is close to one, then the resulting regression coefficient will have a double exponential (Laplace) distribution (slab). If π approximates zero, then the coefficient will approximate zero, too (spike). As a consequence, the aBSS-lasso can be used as a variable selection tool.

In contrast to previous similar implementations for regression models, we see some advantages for the proposed method. In comparison to a discrete mixture model in Equation 8 (Dellaportas et al., 2002; Kuo & Mallick, 1998; Lykou & Ntzoufras, 2013), it allows for a higher flexibility and amount of adaptiveness (Ishwaran & Rao, 2005) as well as a smaller computational burden. Further, using the prior specifications given above, the model can be feasibly implemented in available Bayesian software packages (e.g., in Stan or Jags; Plummer, 2017; Stan Development Team, 2016b), which allows a straightforward use by applied researchers. And, in contrast to the procedure suggested by Ishwaran and Rao (2005) described above, which is an extension of ridge regression that uses a continuous mixture of normal distributions, our method

[3] Simulation results not reported here indicated that the performance of the aBSS-lasso did not depend on the prior specification with Half Cauchy or alternatively with Inverse Gamma distributions.

provides a sharper distribution for the parameter coefficients by using a double exponential (Laplace) prior.

SIMULATION STUDY

In this simulation study, we will now investigate the estimation properties of the proposed aBSS-lasso under different conditions of multicollinearity, reliability, and sample size. We compare its performance to LMS and the UPI approach, which can be viewed as standard procedures for analyzing latent interactions, and to a standard adaptive Bayesian lasso (aB-lasso).

The following complex latent nonlinear model was used in the simulation study:

$$\begin{aligned}\eta_i &= \alpha + \Gamma_1\xi_i + \Gamma_2 vec(\xi_i\xi_i') + \zeta_i \\ &= \alpha + \gamma_{1,1}\xi_{1i} + \gamma_{1,2}\xi_{2i} + \gamma_{1,3}\xi_{3i} + \gamma_{1,4}\xi_{4i} \\ &\quad + \gamma_{2,1}(\xi_{1i})^2 + \gamma_{2,2}\xi_{1i}\xi_{2i} + \gamma_{2,3}\xi_{1i}\xi_{3i} \\ &\quad + \gamma_{2,4}\xi_{1i}\xi_{4i} + \gamma_{2,5}(\xi_{2i})^2 + \gamma_{2,6}\xi_{2i}\xi_{3i} \\ &\quad + \gamma_{2,7}\xi_{2i}\xi_{4i} + \gamma_{2,8}(\xi_{3i})^2 + \gamma_{2,9}\xi_{3i}\xi_{4i} \\ &\quad + \gamma_{2,10}(\xi_{4i})^2 + \zeta_i. \end{aligned} \tag{23}$$

The four predictor variables $(\xi_1, \ldots, \xi_4)$ were generated as standard multivariate normally distributed variables with correlations of $\rho = .2, .4, .6,$ or $.8$; the variance of the (normal) residual ζ was chosen such that η had unit variance and α was set to zero in the population. For data generation, we set the linear effects to $\gamma_{1,1} = \gamma_{1,2} = .3$ and $\gamma_{1,3} = \gamma_{1,4} = 0$; for the nonlinear effects, an interaction between ξ_1, ξ_2 and a quadratic effect for ξ_4 with $\gamma_{2,2} = \gamma_{2,10} = .2$ were included. The size of the nonlinear effects was in line with typical effect sizes reported in Chaplin (1991, 2007). All remaining nonlinear effects were set to zero.

For each latent variable, data for three indicator variables were generated by

$$\mathbf{y}_i = \Lambda_y\eta_i + \in_i \text{ and } \mathbf{x}_i = \Lambda_x\xi_i + \delta_i, \tag{24}$$

where the factor loading matrix Λ_x had simple structure (i.e., each indicator variable only loaded on one latent variable) and Λ_y was a vector for the factor loadings on η. The factor loadings were set to 1. We selected two conditions of reliability (low vs. high) by setting the variances of the residual variables either to 0.25 or 1, which resulted in a reliability for the indicator variables of 0.8 or 0.5, respectively. Indicator variables for the predictor variables were centered within each data set before analysis. Data were generated for sample sizes of $N = 200, 400$ and 800. With the four different conditions of multicollinearity ($\rho = .2, .4, .6, .8$) and two different conditions of reliability (low vs. high), this resulted in 24 data conditions. The complete model is illustrated in Figure 1.

Model Estimation

LMS

LMS was used in its implementation in Mplus. For estimation, a Gauss–Hermite numerical integration with

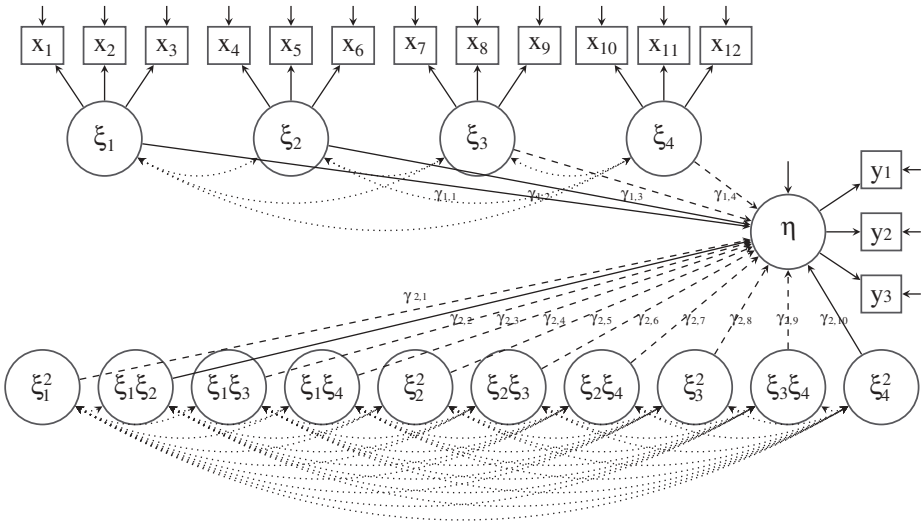


FIGURE 1 Path diagram for the model used in the simulation study. Dashed lines indicate paths that were set to zero in the population model (but estimated freely in the analysis model). Correlations between product terms and the factors $\xi_1, \ldots, \xi_4$ are zero for centered normal variables. Measurement models for the product terms were only used for the UPI approach (and included additional 62 residual covariances between the product indicators).

10 adaptive integration points per dimension was used. This number is lower than that suggested by Klein and Moosbrugger (2000); however, due to computational limitations, this reduction of the integration points was necessary.

Note that the performance of LMS was strongly tied to the specifics of the numerical integration used here. Simulation results not reported here that used a Monte Carlo integration method (with a standard 500 iterations) resulted in very low convergence rates (below 20% for higher correlation conditions). Similar results were obtained for a Gauss–Hermite integration without adaptive integration points (which was suggested by Klein & Moosbrugger, 2000); here, again, convergence rates were very low.

UPI Approach

For the specification of the UPI approach, three product indicators were formed prior to the analysis and were used as indicator variables for the 10 latent product terms. Additionally, several residual error covariances needed to be specified (for details, see Kelava & Brandt, 2009) . For the model as specified in this scenario, a total of 62 residual covariances were necessary. An inspection of the results indicated that these covariances were substantial and different from zero across all conditions. The UPI approach was implemented in Mplus, using a robust MLR estimator with sandwich-type standard errors (White, 1982; Yuan & Bentler, 2000) that were robust to the nonnormality that was induced by including the product indicators.

Lasso Approaches

We selected two implementations of the Bayesian lasso: the aB-lasso and the aBSS-lasso. The aB-lasso follows the model as described for adaptive Bayesian regression models by Leng et al. (2014), using a double exponential (Laplace) prior for the regression coefficients. The aBSS-lasso extends this specification by using a spike-and-slab-prior. For both implementations, we used weakly informative priors for all model parameters except of the latent regression coefficients (see details in Supplement 3, see Footnote 1). For the regression coefficients, let $\gamma = (\gamma_{1,1}, \ldots, \gamma_{1,4}, \gamma_{2,1}, \ldots, \gamma_{2,10})$ include all linear and nonlinear effects. For the aB-lasso, double exponential priors were specified as

$$\gamma_u \sim dexp(0, 1/(\sigma_\zeta \lambda_u)) \qquad \lambda_u^2 \sim C^+(0, 2.5) \tag{25}$$

with $u = 1 \ldots 14$ indicating the uth element of γ, and the residual standard deviation (σ_ζ) of η with a prior distribution of $\sigma_\zeta \sim C^+(0, 2.5)$ (cf. Park & Casella, 2008). For the aBSS-lasso, the following priors were specified:

$$\gamma_u = \pi_u \gamma_u^* \qquad \pi_u \sim Beta(.5, .5)$$

$$\gamma_u^* \sim dexp(0, 1/(\sigma_\zeta \lambda_u)) \qquad \lambda_u^2 \sim C^+(0, 2.5) \tag{26}$$

with $Beta(.5, .5)$ as an uninformative Jeffrey's prior.

The lasso approaches were implemented in Stan (Carpenter et al., 2017; Stan Development Team, 2016b) using the Rstan package (Stan Development Team, 2016a) in R (R Core Team, 2015). Four chains with each 2000 iterations were used for estimation. The first 1000 iterations were discarded as burn-in iterations. Stan uses a Hamiltonian Monte Carlo algorithm that converges faster than other Markov Chain Monte Carlo (MCMC) algorithms such as the Gibb's sampler (as they are implemented in jags or openbugs). Hence, a smaller number of iterations is sufficient to ensure convergence (Betancourt & Girolami, 2015; Neal, 2011). Convergence was checked in detail for random samples across all conditions using trace and density plots as well as the Rhat statistic. Furthermore, convergence was checked for each data set by investigating the Rhat statistic and selecting samples only if $Rhat < 1.1$ for all model parameters (Gelman & Rubin, 1992).

RESULTS

In order to provide an informative overview, results for different parameter "categories" were summarized based on their population values: linear nonzero effects ($\gamma_1 \neq 0$, i.e.,$\gamma_{1,1} = \gamma_{1,2} = .3$), nonlinear nonzero effects ($\gamma_2 \neq 0$, i.e.,$\gamma_{2,2} = \gamma_{2,10} = .2$), and the respective linear and nonlinear zero effects ($\gamma_1 = 0$, $\gamma_2 = 0$). Results for parameters within each of these category were very similar, and averaging the results did not lead to either a substantial reduction of information or a distorted presentation of the findings. Tables for the results can be found in Supplement 3 (see Footnote 1).

We present results on convergence rates, the (absolute) parameter bias and coverage rates, and then on the root mean squared errors (RMSE), which is a measure of parameter accuracy. Coverage rates were based on the 95% probability intervals for the lasso approaches. For the frequentist approaches (LMS and UPI), confidence intervals were calculated based on the estimated standard errors.

Convergence Rates

Convergence rates were generally high with average rates of 98.3% for the aBSS-lasso, 99.7% for the aB-lasso, 99.6% for LMS, and 91.7% for the UPI approach. The UPI approach produced lower convergence rates for small sample sizes and low reliability (60.0–79.0%) as well as for high multicollinearity and low reliability (82.0% for $N = 400$ and 79.0% for $N = 800$).

Parameter Bias

Figure 2 shows the results of the parameter bias. For the linear effects (γ_1, left panels), the bias for all approaches lay between ± 0.04 except for the condition with small sample size, low reliability, and high multicollinearity: here, the UPI approach produced a bias of -0.09 for linear nonzero effects and $+0.10$ for the linear zero effects. Some bias was also found for the lasso approaches (between -0.05 and 0.06) under this condition. For the nonlinear effects that were zero in the population ($\gamma_2 = 0$, right bottom panel), parameters were mostly unbiased (with values between -0.03 and 0.04) across all conditions.

For the nonzero coefficients ($\gamma_2 \neq 0$, right top panel), differences across the estimators were found. The UPI approach produced the largest bias, ranging from -0.18 to 0.16 for low-reliability conditions and between -0.01 to 0.13 for high-reliability conditions. The bias was more prominent in the low-reliability conditions, where even for large sample sizes ($N = 800$) a bias of -0.15 was observed for the strong multicollinearity condition. LMS produced relatively unbiased estimates except for some marginal bias for small sample sizes ($N = 200$), high multicollinearity and low reliability (with a bias up to –0.03).

Both lasso implementations (aBSS and aB) performed similarly. For low-reliability conditions, the negative bias increased with multicollinearity up to -0.13; for high-reliability conditions, the bias increased up to -0.10. The bias was slightly stronger for the aBSS-lasso compared to the aB-lasso. The negative bias was in line with the model specification and was expected: small effects were drawn to zero because the prior was chosen with a zero mean, pulling effects closer to zero in order to produce a sparse parameter vector. The bias decreased with increasing sample size.

Coverage

Figure 3 illustrates the 95% coverage rates. For the aBSS-lasso, coverage rates lay between 0.92 and 0.98 for nonzero effects and above 0.98 for zero effects. For the aB-lasso similar coverage rates were found for nonzero effects (0.92 to 0.99) but slightly smaller rates compared to the aBSS lasso for zero effects (0.95 to 1.00). Coverage rates were very similar for low- and high-reliability conditions as well as for different sample sizes or multicollinearity conditions. The bias found for the parameter estimates of the lasso approaches did not severely affect the coverage rates.

For LMS, coverage rates lay between 0.92 and 0.98; there was virtually no difference between zero and nonzero effects and only minor differences across conditions. For the UPI approach, differences could be observed between high- and low-reliability conditions. For low reliability, coverage rates lay between 0.96 and 1.00, and for high reliability, coverage rates lay between 0.90 and 1.00; they were higher for low-reliability conditions and increased with the amount of multicollinearity.

RMSE

As can be seen from Figure 4, the lasso approaches produced more accurate estimates compared to the other approaches particularly in situations with high multicollinearity and low reliability. The RMSE of the lasso approaches were relatively unaffected by increasing multicollinearity or sample size. It lay between 0.02 and 0.19 for the aBSS-lasso and between 0.03 and 0.21 for the aB-lasso. The RMSE was slightly larger for low-reliability condition (0.03–0.19 for the aBSS-lasso and 0.05–0.21 for the aB-lasso) than for the high-reliability conditions (0.02–0.14 and 0.03–0.17, respectively).

The RMSE of both LMS and the UPI approach were strongly affected by (low) reliability, increasing multicollinearity, and (small) sample size. For LMS, the RMSE increased more strongly for nonlinear effects compared to linear effects; there was virtually no difference between zero and nonzero effects. For low reliability, the RMSE increased from 0.05 to 0.83, and for high reliability, the RMSE increased from 0.03 to 0.29. For the UPI approach, the RMSE increased up to 1.18 (and 1.17) for low (and high)-reliability conditions.

Finally, a comparison of the relative accuracy between approaches showed that the aBSS-lasso produced the most accurate estimates particularly for zero parameters (see Figure 5). The relative accuracy was defined as the RMSE of the respective approach divided by the RMSE of the aBSS-lasso under each condition. Values above 1 indicate a smaller RMSE for the aBSS-lasso; values below 1 indicate a smaller RMSE for the respective approach.

Compared to the aB-lasso, a similar accuracy could be found for nonzero effects (ratios lay between 0.90 and 1.36); for zero effects, the aBSS-lasso was more accurate with ratios between 1.25 and 2.00, which indicated that the RMSE was up to twice as large for the aB-lasso compared to the aBSS-lasso. Differences across sample sizes, multicollinearity, and reliability were marginal.

For LMS, accuracy was similar only for the linear nonzero effects. For the linear zero effects, the accuracy was about 1.8 times higher for the aBSS-lasso compared to LMS. For the nonlinear effects, the advantage of the aBSS-lasso increased with the amount of multicollinearity. It was more pronounced for low-reliability conditions with ratios up to 8.28 compared to high-reliability conditions with ratios up to 3.30.

A similar result was found for the UPI approach. Here, the accuracy of the aBSS-lasso was up to 12.82 (or 11.97) times higher than for the UPI approach for low (or high)-reliability conditions. A comparable accuracy for the two approaches could only be found for linear nonzero effects under the condition of high reliability and large sample size ($N = 800$).

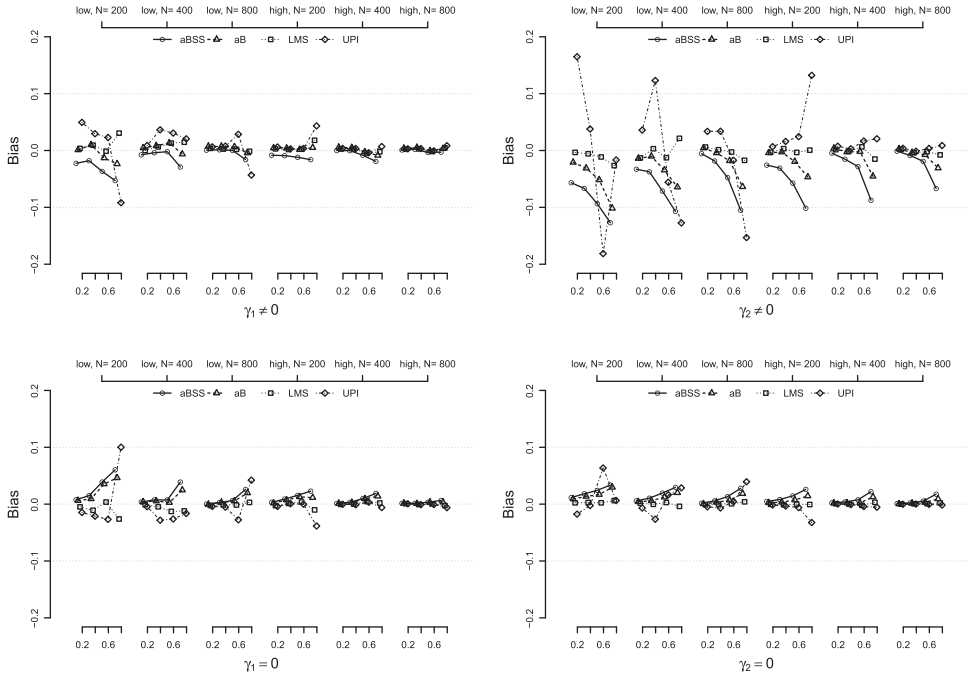


FIGURE 2 Average parameter bias for zero ($\gamma_1 = 0, \gamma_2 = 0$) and nonzero ($\gamma_1 \neq 0, \gamma_2 \neq 0$) linear and nonlinear parameters.

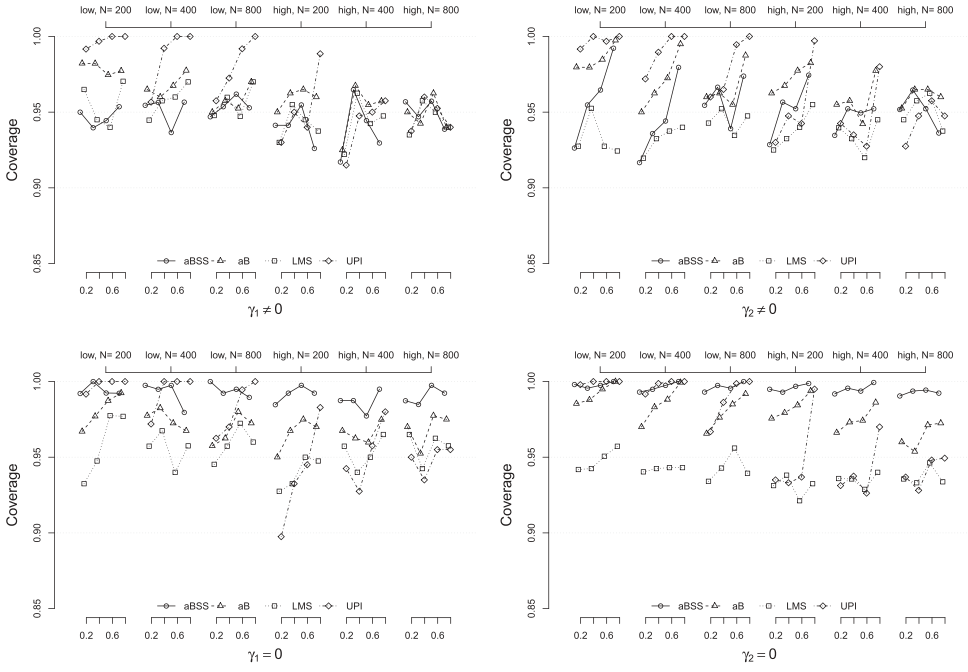


FIGURE 3 Average 95% coverage rates for zero ($\gamma_1 = 0, \gamma_2 = 0$) and nonzero ($\gamma_1 \neq 0, \gamma_2 \neq 0$) linear and nonlinear parameters.

DISCUSSION

In this article, we presented the aBSS-lasso, which is an adaptive Bayesian lasso model with spike-and-slab priors for SEM with latent linear and nonlinear interaction and quadratic effects. Simulation results indicated that the advantages of the aBSS-lasso compared to standard approaches (LMS and UPI) were most pronounced for smaller sample sizes, low reliability, and increasing multicollinearity. This study also provided

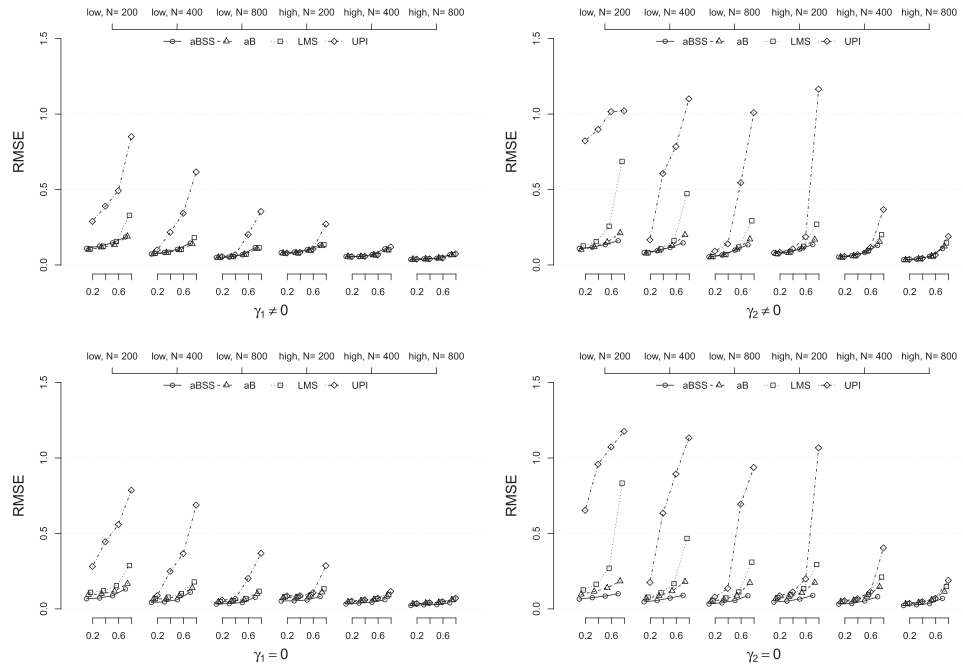


FIGURE 4 Average RMSE for zero ($\gamma_1 = 0, \gamma_2 = 0$) and nonzero ($\gamma_1 \neq 0, \gamma_2 \neq 0$) linear and nonlinear parameters.

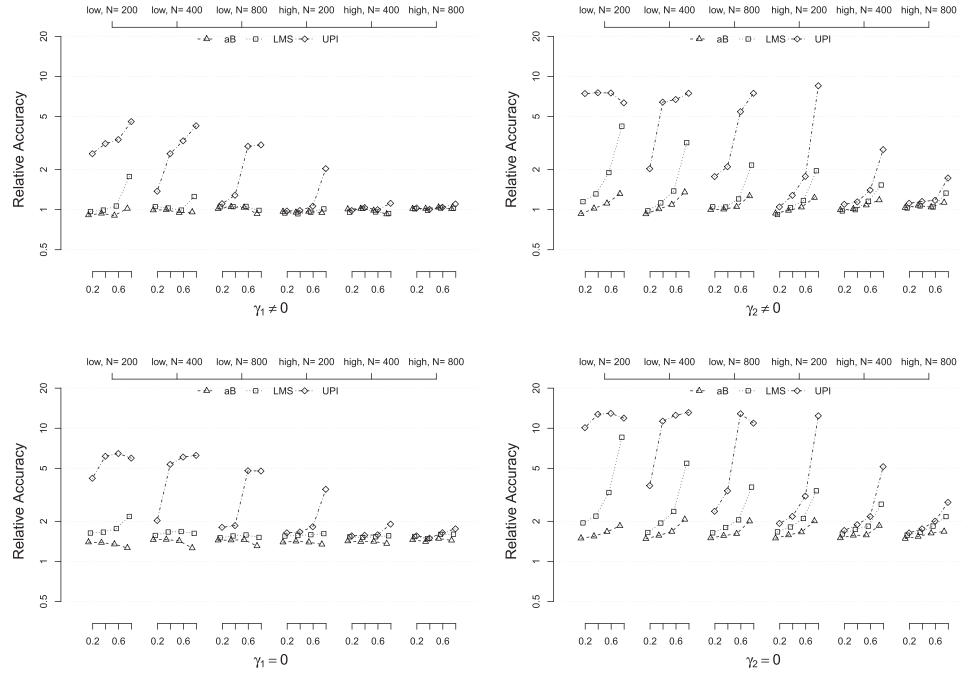


FIGURE 5 Relative accuracy for zero ($\gamma_1 = 0, \gamma_2 = 0$) and nonzero ($\gamma_1 \neq 0, \gamma_2 \neq 0$) linear and nonlinear parameters. The relative accuracy was calculated by dividing the RMSE of the respective approach and the one of the aBSS-lasso. Note that the *y* axis has a logarithmic scale.

evidence that this particular type of Bayesian lasso approach produces more accurate results in the (nonlinear) SEM framework compared to another recent adaptive Bayesian lasso approach (i.e., aB-lasso) that uses a double exponential prior but not a spike-and-slab prior (e.g., Feng et al., 2017).

The main advantage of the aBSS-lasso compared to standard approaches was that accuracy did not depend as strongly

on the correlation between predictor variables. The UPI approach showed less favorable estimation properties particularly in situations with low reliability or high multicollinearity that resulted in an RMSE that was up to 12 times larger compared to the aBSS-lasso. The reason for the suboptimal performance of the UPI approach lies in its conceptualization: it needs product indicators. If several latent product terms are estimated simultaneously, a substantive amount of multicollinearity between these product indicators is induced because they are created from the same indicators (e.g., X_1 was included in X_1X_4 for a latent interaction $\xi_1\xi_2$ and in X_1X_7 for $\xi_1\xi_3$ in the simulation study). This multicollinearity is taken into account by including several residual covariances, which in turn makes the model less parsimonious with increasing numbers of latent product terms in the model (for technical details, see Kelava & Brandt, 2009).

LMS' performance was impaired mostly by the amount of multicollinearity, which resulted in an RMSE as high as about eight times the one of the aBSS-lasso. Furthermore, models that are more complex than those investigated in the simulation study might lead to computational limitations for LMS because the computational burden increases exponentially with the number of latent product terms (and resulting dimensions of integration). The aBSS-lasso as a Bayesian method will of course result in more time consuming analyses but will still provide results.

The theoretical advantages of the aBSS-lasso compared to the aB-lasso are that the shrinkage for small effects is more effective (i.e., higher accuracy) and that the method allows to retrieve more of its original intention as a selection tool. In the simulation study, we showed that both implementations provide similar results in some situations. Practically relevant advantages for the aBSS-lasso were found for effects that are zero in the population. Here, the aBSS-lasso produced more accurate results as indicated by up to a twice as small an RMSE as compared to the aB-lasso.

Limitations

As with any simulation study, the number of scenarios investigated was limited. When using complex nonlinear models, it is also necessary to consider the processing time that increases with sample size and model complexity. Both LMS and the lasso implementations are time-consuming methods. In this simulation, we focused on normally distributed data and we did not include effects due to clustering (i.e., multilevel structures). Extensions of the model formulation are straightforward to implement, for example, by using mixture models (Kelava et al., 2014) or multilevel SEM. However, the simulation included relevant conditions of reliability, multicollinearity, and sample size, which revealed substantial differences between the methods investigated.

One finding in the simulation study was a negative bias of the lasso implementations for effects that were different from zero in the population. This bias was to be expected due to the prior formulation. The main scope of the spike-and-slab prior (or the double exponential prior) is to shrink small parameters to zero in order to produce results that have a "simple structure" and illustrates the method's variable selection property. Coverage rates were not impaired and lay well above 90%, which implied that the bias did not produce results that would lead to wrong substantive conclusions.

Comparison to Other Shrinkage Methods

Bayesian shrinkage priors have recently received more attention, resulting in several different priors (Polson & Scott, 2011). Spike-and-slab priors with categorical mixing variables are considered to be two-group models because model parameters are discrete mixtures of a point-mass and another distribution. They are sometimes considered the gold standard for shrinkage methods (Piironen & Vehtari, 2017). One of the disadvantages of these discrete models is that they result in high-dimensional discrete parameter spaces that are computationally demanding (Bhadra et al., 2017). The model formulation used here for the aBSS-lasso and the model in Ishwaran and Rao (2005) overcomes the limitation of previous spike-and-slab priors by using a continuous mixture.

The second group of Bayesian shrinkage priors are considered one-group models or continuous scale mixtures (Bhadra et al., 2017). This group includes many different shrinkage methods such as the horseshoe prior (Carvalho, Polson, & Scott, 2010), the horseshoe+ prior (Bhadra et al., 2017), the normal gamma prior (Griffin & Brown, 2010), and others (e.g., Armagan, Dunson, & Lee, 2013; Armagan, Clyde, & Dunson, 2011; Bhattacharya, Pati, Pillai, & Dunson, 2015; Zhang, Reich, & Bondell, 2016). The scale-mixture priors have been developed particularly for sparse regression models that include, for example, 1000 predictor variables (e.g., in genome research). One of their advantages is that they are fast and easy to implement (e.g., in Stan). However, one of the disadvantages, at least for the horseshoe prior, is that the posterior distribution can be multimodal if predictor variables are highly correlated (Piironen & Vehtari, 2017).

The use of a continuous mixture in the aBSS-lasso formulation leads to a closer resemblance of spike-and-slab priors and scale mixture priors. Future conceptual research and simulations are needed to compare the different shrinkage priors for linear or nonlinear SEM.[4] This comparison, though, lies beyond the scope of this article.

[4] Additional simulation results not reported here indicated that the horseshoe prior has very similar characteristics to the aBSS-lasso with regard to bias and accuracy; however, convergence rates were lower with an average of 74%. They improved with sample size from 69% for $N = 200$ to 80% for $N = 800$.

Practical Implications and Future Directions

Shrinkage methods, such as the lasso, only have advantages in more complex models and when the actual parameter vector is sparse (i.e., many zeros). From reading the literature, we know that there are many situations in social, behavior, and education sciences that involve these conditions. First, when more complex theories (such as the expectancy-value theory) are investigated, it is meaningful to include these relationships and hypotheses into a single model. If relevant effects are omitted, the remaining effects may have spurious estimates (Klein et al., 2009). Theories in many areas become more complex in order to describe human behavior.

Second, large-scale data sets open up the possibility to investigate relevant hypotheses under consideration of many covariates. This possibility allows one to formulate more general models, which in turn allows researchers to generalize effects across different subpopulations described by the covariates. One practical example is DIF detection. Following the model formulation put forth by Bauer (2017), covariates can be included to detect DIF in factor loadings using products between factors and covariates. We expect that the aBSS-lasso can provide an important improvement in estimation and model selection in situations where many covariates and items are available.

Third, the structure of available data sets becomes more complex, for example, when data are collected from schools (such as in TIMSS or PISA), or in multicenter treatment studies. Frameworks that can account for such clustered structures have been developed in general (Asparouhov, Hamaker, & Muthén, 2017; Kelava & Brandt, 2014; Muthén & Asparouhov, 2009). While some researchers warn against the formulation of overly complex models (e.g., Preacher, Zhang, & Zyphur, 2016), other recent work on misspecified SEM showed that overly restrictive models often lead to severe bias (e.g., Bollen, Kolenikov, & Bauldry, 2014). The aBSS-lasso may help to solve this dilemma because it performs well when models are complex, but provides sparse parameter estimates that allow researchers to formulate a more parsimonious model for subsequent analyses. Particularly with regard to lacking model fit for nonlinear models (Gerhard et al., 2014), it is important to provide methods that can fit more complex models that resemble saturated models, which potentially include the correct model as a submodel.

Fourth, semiparametric modeling using latent variables has recently become more prominent (e.g., Feng et al., 2017; Kelava & Brandt, 2014; Song et al., 2013). These models substantially increase computational demands because their formulation (e.g., using basis expansion for splines) results in multiple latent variables for each predictor included in the model. The applicability of shrinkage methods to semiparametric modeling can improve model estimation and interpretation, for example, by smoothing the extracted nonlinear function. Some shrinkage methods, such as an adaptive lasso, have been used (Guo et al., 2012), but more recent methods such as the spike-and-slab prior might lead to further improvement.

Final Remarks

The aBSS-lasso produced more accurate results in situations of high multicollinearity and low reliability compared to other approaches when analyzing many effects simultaneously. When analyzing complex theoretically implied associations, such as in studying motivation, or large-scale data sets with many variables, this method can be applied because it can accommodate many effects simultaneously. Due to its spike-and-slab prior implementation, the aBSS-lasso estimator can be used to select relevant linear and nonlinear effects more accurately than other methods. In situations with only few effects and predictor variables (such as a single interaction effect), it can be assumed that there are no advantages from this method compared to traditional approaches such as LMS.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft [KE 1664/1-2].

REFERENCES

Armagan, A., Clyde, M., & Dunson, D. B. (2011). Generalized beta mixtures of Gaussians. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 24, pp. 523–531). Red Hook, NY: Curran Associates, Inc.

Armagan, A., Dunson, D. B., & Lee, J. (2013). Generalized double pareto shrinkage. *Statistica Sinica*, *23*, 119–143. doi:10.5705/ss.2011.048

Asparouhov, T., Hamaker, E. L., & Muthén, B. O. (2017). Dynamic latent class analysis. *Structural Equation Modeling*, *24*, 257–269. doi:10.1080/10705511.2016.1253479

Atchade, Y. F. (2011). A computational framework for empirical Bayes inference. *Statistics and Computing*, *21*, 463–473. doi:10.1007/s11222-010-9182-3

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*, 507–526. doi:10.1037/met0000077

Betancourt, M., & Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. In S. K. Upadhyay, U. Singh, D. K. Dey, & A. Loganathan (Eds.), *Current trends in Bayesian methodology with applications* (pp. 79–102). Boca Raton, FL: Chapman & Hall/CRC.

Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, *12*, 1105–1131. doi:10.1214/16-BA1028

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet-Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, *110*, 1479–1490. doi:10.1080/01621459.2014.960967

Bollen, K. A., Kolenikov, S., & Bauldry, S. (2014). Model-implied instrumental variable – Generalized method of moments (MIIV-GMM) estimators for latent variable models. *Psychometrika*, *79*, 20–50. doi:10.1007/s11336-013-9335-3

Brandt, H., Kelava, A., & Klein, A. G. (2014). A simulation study comparing recent approaches for the estimation of nonlinear effects in SEM

under the condition of non-normality. *Structural Equation Modeling*, *21*, 181–195. doi:10.1080/10705511.2014.882660

Brandt, H., Umbach, N., & Kelava, A. (2015). The standardization of linear and nonlinear effects in direct and indirect applications of structural equation mixture models for normal and nonnormal data. *Frontiers in Psychology*, *6* (1813), 1–13. doi:10.3389/fpsyg.2015.01813

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. doi:10.18637/jss.v076.i01

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

Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics*, *2*, 485–500. doi:10.1093/biostatistics/2.4.485

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: A comparison of four approaches. *Multivariate Behavioral Research*, *47*, 840–876. doi:10.1080/00273171.2012.732901

Chaplin, W. (1991). The next generation of moderator research in personality psychology. *Journal of Personality*, *59*, 143–178. doi:10.1111/jopy.1991.59.issue-2

Chaplin, W. (2007). Moderator and mediator models in personality research: A basic introduction. In R. W. Robins, C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 602–632). New York, NY: Guilford.

Dakanalis, A., Timko, C. A., Zanetti, M. A., Rinaldi, L., Prunas, A., Carrà, G., … Clerici, M. (2014). Attachment insecurities, maladaptive perfectionism, and eating disorder symptoms: A latent mediated and moderated structural equation modeling analysis across diagnostic groups. *Psychiatry Research*, *215*, 176–184. doi:10.1016/j.psychres.2013.10.039

Dellaportas, P., Forster, J. J., & Ntzoufras. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, *12*, 27–36. doi:10.1023/A:1013164120801

Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, *106*, 569. doi:10.1037/a0035504

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives. Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: Freeman.

Fahrmeir, L., Kneib, T., & Konrath, S. (2010). Bayesian regularisation in structured additive regression: A unifying perspective on shrinkage, smoothing and predictor selection. *Statistics and Computing*, *20*, 203–219. doi:10.1007/s11222-009-9158-3

Feng, X.-N., Wang, G.-C., Wang, Y.-F., & Song, X.-Y. (2015). Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Statistics in Medicine*, *34*, 1527–1547. doi:10.1002/sim.6410

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2015). Bayesian adaptive lasso for ordinal regression with latent variables. *Sociological Methods & Research, (Online)*, 1–28. doi:10.1177/0049124115610349

Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling, 24*, 341-358. doi:10.1080/10705511.2016.1257353

Frank, I., & Friedman, J. (1993). A statistical view of some chemometrics regression tools. *Technometrics : A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, *35*, 109–148. doi:10.2307/1269656

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, *3*, 235–247. doi:10.1037/1082-989X.2.3.235

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534. doi:10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511. Retrieved from www.jstor.org/stable/2246093

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

Gerhard, C., Büchner, R. D., Klein, A. G., & Schermelleh-Engel, K. (2014). A fit index to assess model fit and detect omitted terms in nonlinear sem. *Structural Equation Modeling*, *24*, 414–427. doi:10.1080/10705511.2016.12689239

Gerhard, C., Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., Gäde, J., & Brandt, H. (2014). On the performance of likelihood-based difference tests in nonlinear structural equation models. *Structural Equation Modeling*, *22*, 276–287. doi:10.1080/10705511.2014.935752

Griffin, J. E., & Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*, 171–188. doi:10.1214/10-BA507

Griffin, J. E., & Brown, P. J. (2011). Bayesian hyper-lassos with non-convex penalization. *Australian and New Zealand Journal of Statistics*, *53*, 423–442. doi:10.1111/j.1467-842X.2011.00641.x

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences*, *37*, 161–168. doi:10.1016/j.lindif.2015.01.008

Guo, R., Zhu, H., Chow, S.-M., & Ibrahim, J. G. (2012). Bayesian lasso for semiparametric structural equation models. *Biometrics*, *68*, 567–577. doi:10.1111/j.1541-0420.2012.01751.x

Hans, C. (2009). Bayesian lasso regression. *Biometrika*, *96*, 835–845. doi:10.1093/biomet/asp047

Hao, S., Hong, W., Xu, H., Zhou, L., & Xie, Z. (2015). Relationship between resilience, stress and burnout among civil servants in Beijing, china: Mediating and moderating effect analysis. *Personality and Individual Differences*, *83*, 65–71. doi:10.1016/j.paid.2015.03.048

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

Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *The Annals of Statistics*, *33*, 730–773. Retrieved from www.jstor.org/stable/3448605

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. doi:10.1080/10705511.2016.1154793

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–87). Mahwah, NJ: Erlbaum.

Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, *16*, 123–131.

Kelava, A., & Brandt, H. (2014). A general nonlinear multilevel structural equation mixture model. *Frontiers in Quantitative Psychology and Measurement*, *5*, 748. doi:10.3389/fpsyg.2014.00748

Kelava, A., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for non-normally distributed latent predictor variables. *Structural Equation Modeling*, *21*, 468–481. doi:10.1080/10705511.2014.915379

Kelava, A., Werner, C., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., … West, S. G. (2011). Advanced nonlinear structural equation modeling: Theoretical properties and empirical application of the distribution-analytic LMS and QML estimators. *Structural Equation Modeling*, *18*, 465–491. doi:10.1080/10705511.2011.582408

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi:10.1007/BF02296338

Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–674. doi:10.1080/00273170701710205

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Assessing spurious interaction effects. In T. Teo & M. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 13–28). Rotterdam, NL: Sense Publishers.

Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhy¯A (B)*. *60*, 65–81. Retrieved from www.jstor.org/stable/25053023

Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, *5*, 369–412. doi:10.1214/10-BA607

Lapan, R. T., & Jingelski, J. (1992). Circumscribing vocational aspirations in junior secondary school. *Journal of Counseling Psychology*, *39*, 81–90. doi:10.1037/0022-0167.39.1.81

Leng, C., Tran, M.-N., & Nott, D. (2014). Bayesian adaptive lasso. *Annals of the Institute of Statistical Mathematics*, *66*, 221–244. doi:10.1007/s10463-013-0429-6

Lykou, A., & Ntzoufras, I. (2013). On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, *23*, 361–390. doi:10.1007/s11222-012-9316-x

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300. doi:10.1037/1082-989X.9.3.275

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B*, *70*, 53–71. doi:10.1111/j.1467-9868.2007.00627.x

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

Mullis, I., & Martin, M. E. (2013). *TIMSS 2015 assessment frameworks*. TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo* (pp. 113–162). Boca Raton, FL: Chapman & Hall/CRC.

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *482*, 681–686. doi: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. doi:10.1214/17-EJS1337SI

Plummer, M. (2017). *Jags version 4.3.0 user manual [Computer software manual]*. Retrieved from https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/

Polson, N. G., & Scott, J. G. (2011). Shrink globally, act locally: Sparse Bayesian regularization and prediction. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian statistics* (Vol. 9, pp. 501–538). New York, NY: Oxford University Press.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, *21*, 189–205. doi:10.1037/met0000052

R Core Team. (2015). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Retrieved from https://www.R-project.org

Scalco, M. D., Colder, C. R., Hawk, L. W., Jr, Read, J. P., Wieczorek, W. F., & Lengua, L. J. (2014). Internalizing and externalizing problem behavior and early adolescent substance use: A test of a latent variable interaction and conditional indirect effects. *Psychology of Addictive Behaviors*, *28*, 828. doi:10.1037/a0035805

Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.

Song, X.-Y., Li, Z.-H., Cai, J.-H., & Ip, E. H.-S. (2013). A Bayesian approach for generalized semiparametric structural equation models. *Psychometrika*, *78*, 624–647. doi:10.1007/s11336-013-9323-7

Stan Development Team. (2016a). *RStan: The R interface to Stan*. Retrieved from http://mc-stan.org (R package version 2.14.1)

Stan Development Team. (2016b). *Stan modeling language: User's guide and reference manual*. Retrieved from http://mc-stan.org (Version 2.14.0)

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*. *58*, 267–288. Retrieved from http://www.jstor.org/stable/2346178

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*, 763–777. doi:10.1037/a0027470

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*, 47–63. doi:10.1348/000711003321645331

Wang, H., Li, G., & Tsai, C. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *69*, 63–78. doi:10.1111/j.1467-9868.2007.00577.x

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*, 1556–1574. doi:10.1111/j.1467-8624.2004.00757.x

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica : Journal of the Econometric Society*, *50*, 1–25. doi:10.2307/1912526

Wigfield, A., & Cambria, J. (2010a). Expectancy-value theory: Retrospective and prospective. In T. Urdan & S. Karabenick (Eds.), *The decard ahead: Applications and contexts of motivation and achievement* (Vol. Advances I, pp. 35–70). Bingley, UK: Emerald Group Publishing Limited. doi:10.1108/S0749-7423(2010)000016A005

Wigfield, A., & Cambria, J. (2010b). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*, 1–35. doi:10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. doi:10.1006/ceps.1999.1015

Xu, X., & Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis*, *4*, 909–936. doi:10.1214/14-BA929

Yuan, K. H., & Bentler, P. M. (2000). Three likelihoodbased methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200. doi:10.1111/0081-1750.00078

Yuan, M., & Lin, Y. (2005). Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, *100*, 1215–1225. doi:10.1198/016214505000000367

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B*, *68*, 49–67. doi:10.1111/j.1467-9868.2005.00532.x

Zhang, Y., Reich, B. J., & Bondell, H. D. (2016). *High dimensional linear regression via the r2-d2 shrinkage prior*. arXiv preprint. arXiv:1609.00046.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*, 1418–1429. doi:10.1198/016214506000000735

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Ser. B*, *67*, 301–320. doi:10.1111/j.1467-9868.2005.00503.x