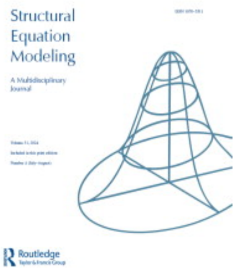
Routledge
Taylor & Francis Group



Structural Equation Modeling: A Multidisciplinary Journal

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

Research Design and Model Estimation Under the Partially Confirmatory Latent Variable Modeling Framework with Multi-Univariate Bayesian Lassos

Jinsong Chen & Yifan Zhang

To cite this article: Jinsong Chen & Yifan Zhang (04 Sep 2024): Research Design and Model Estimation Under the Partially Confirmatory Latent Variable Modeling Framework with Multi-Univariate Bayesian Lassos, Structural Equation Modeling: A Multidisciplinary Journal, DOI: 10.1080/10705511.2024.2392618

To link to this article: https://doi.org/10.1080/10705511.2024.2392618

Published online: 04 Sep 2024.

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2024.2392618

Research Design and Model Estimation Under the Partially Confirmatory Latent Variable Modeling Framework with Multi-Univariate Bayesian Lassos

Jinsong Chen and Yifan Zhang

The University of Hong Kong

ABSTRACT

This research builds upon existing developments of the partially confirmatory approach by introducing predictors and regularizations to two additional parameter matrices: structural and differential coefficients. The outcome is a comprehensive framework called partially confirmatory latent variable modeling (PCLVM), where researchers can apply different regularizations to four parameter matrices individually or collectively, and in full or in part. With PCLVM, applied researchers can design a variety of research studies for different purposes, depending on the combinations of different regularizations. It employs a mixed estimation algorithm combining univariate and multivariate Bayesian Lassos for measurement- and structural-level regularizations with or without correlated residuals. The attractiveness of the proposed framework was demonstrated through a variety of typical cases that can be readily estimated and widely encountered in practice. Simulation studies and real-life data analysis were adopted to showcase the performance and versatility of PCLVM and its comparisons with exploratory structural equation modeling.

KEYWORDS

Bayesian lasso; latent variable; measurement level; partially confirmatory; structural level

1. Introduction

The latent variable modeling (LVM) framework is a powerful family of multivariate statistical techniques that provides researchers with different approaches to address major research processes. LVM covers a wide range of models, including factor analysis (FA), item response theory models, latent class analysis, and structural equation modeling (SEM) with latent variables. The flexibility of LVM allows researchers to fully exploit a variety of latent-variable concepts such as factors, latent traits, abilities, personalities, attitudes, and so on. As a result, LVM is widely used across many social and behavioral disciplines such as education, psychology, sociology, economics, and business, to name a few.

Within the LVM framework, there are two typical research approaches: exploratory and confirmatory. The exploratory approach is purely data-driven with little substantive knowledge available, while the confirmatory approach is theory-driven and requires strong substantive knowledge. In FA, the exploratory approach is known as exploratory FA (EFA), and the confirmatory approach is known as confirmatory FA (CFA). There are different perspectives to interpret the two approaches. From the perspective of data analysis, EFA can be referred to as unrestricted FA, with CFA as restricted FA or a constrained version of EFA. However, this perspective ignores substantial differences between the two approaches, like if the number of factors is known, if the model is identified, and if parameter estimation needs to be precisely accurate.

From the perspective of research design or scale development, the two approaches can be perceived as two ends of an exploratory-confirmatory continuum covering varying amounts of theoretical or substantive knowledge available. For example, when developing a new measurement scale, researchers may have strong knowledge about what a subset of items measures, but not about others. The goal, therefore, is to find a more flexible approach that can cover the continuum more effectively under the LVM framework.

Recent advances in regularization methods offer greater flexibility to incorporate different amounts of substantive knowledge within existing procedures, covering a wider range of the confirmatory-exploratory continuum (e.g., Huang et al., 2017; Jacobucci et al., 2016; Lu et al., 2016). These methods allow researchers to specify a range of prior information, from complete theory-driven models to completely data-driven models, and everything in between. By allowing for greater flexibility, these methods can help researchers develop more accurate models that reflect the complexity of real-world phenomena. Regularization methods penalize the likelihood function with different types of parameter norms for model simplicity. Two typical cases for regression are the ridge (Hoerl & Kennard, 1970) and Lasso[1] (Tibshirani, 1996) estimators, which correspond to the L_2-norm and L_1-norm penalization, respectively. Both

CONTACT Jinsong Chen jinsong.chen@live.com Faculty of Education, The University of Hong Kong, Room 420, 4/F, Meng Wah Complex, Pokfulam Road, Hong Kong.
An earlier version of this work was presented at the 2023 Annual Meeting of the National Council on Measurement in Education.
[1]Lasso is the acronym for least absolute shrinkage and selection operator.

estimators can be solved analytically under the frequentist approach, and interested readers can refer to the literature for details (e.g., Hastie et al., 2009). Corresponding Bayesian regularization methods such as the Bayesian Lasso for regression can be implemented in a hierarchical formulation with appropriate priors (Park & Casella, 2008).

For LVM, both the frequentist regularization (e.g., Jacobucci et al., 2016) and Bayesian regularization (e.g., Chen et al., 2021; Feng et al., 2017; Lu et al., 2016) approaches have been implemented. Compared to the frequentist approach, Bayesian regularization enjoys the benefits of interval estimates and related significance testing, straightforward estimation of the shrinkage parameters, and scalability to complex models that are analytically intractable. Moreover, Bayesian regularization can effectively address different challenges when regularizing the factor loadings and/or local dependence (LD, i.e., correlated residuals).

More recently, a partially confirmatory approach to psychometric model with Bayesian regularization was introduced to tackle a series of psychometric issues (Chen, 2020, 2021b; Chen et al., 2021). It was based on the partially confirmatory factor analysis (PCFA), which can regularize both the loading structure and LD simultaneously for continuous data (Chen et al., 2021). Based on different types of Bayesian Lasso and some constraints, the approach offers a two-step procedure to tackle the two issues in sequence, which can provide greater flexibility for modeling, especially with multiple factors and many items. The approach was also extended to cover exploratory settings with unknown number of factors (Chen, 2021a, 2023). The strength of the approach lies in its power to integrate partial knowledge, model simplicity, interval estimate, and scalability to mixed-response format, missingness, and LD (Chen, 2022). Meanwhile, the potential of the approach is far from exhausted.

This research expands the PCFA approach by incorporating predictors and regularizations on structural parametric matrices. Specifically, it suggests combining PCFA with the multivariate generalized latent variable model (MGLVM; Feng et al., 2017), both of which involve latent variables, the Bayesian approach, and Lasso regularization. In the MGLVM, regularization occurs on the structural level, while the measurement component remains strictly confirmatory. In contrast, regularizations in PCFA take place at the measurement level. By integrating both levels of regularizations, a comprehensive framework, partially confirmatory latent variable modeling (PCLVM) is established. This allows for the regularization of different parameter matrices on both levels, either separately or jointly, as well as fully or partially.

Under PCLVM, one can come up with a variety of research designs that can be used for different purposes, depending on the combinations of different regularizations. While some designs can be addressed using alternative methods like exploratory SEM (ESEM, Asparouhov & Muthén, 2009) or regularized SEM (Jacobucci et al., 2016), many can only be effectively explored through the PCLVM framework. Given that social and behavioral researchers often strive to discern causal relationships or root causes, modeling causation will become more accessible with both sophisticated structural designs (Pearl, 2009) and different regularizations acting as Occam's razor (Blumer et al., 1987). The estimation algorithm is mixed with univariate and multivariate Bayesian Lassos to accommodate measurement- and structural-level regularizations, even with LD. Model estimation is implemented through the Markov chain Monte Carlo (MCMC; Gilks et al., 1996) method with the Gibbs sampler (Casella & George, 1992; Geman & Geman, 1984). The MCMC algorithm will be implemented in an R package and will be made available online to assist future research efforts.

This paper introduces typical designs that can be readily estimated and encountered in practice. The identified designs serve as a starting point for researchers, offering practical and applicable solutions. However, it's worth noting that there are still unexplored designs and potential applications, awaiting further investigation in future research. The Bayesian inference with regularizations is provided, with technical details of the full conditional distributions for the Gibbs sampler. Moreover, we assess the empirical performance of the proposed methodology through simulation studies and utilize response data from the Trends in International Mathematics and Science Study to demonstrate real-life data analysis, accompanying with comparisons to the ESEM approach.

2. General Formulation

Assume that a set of observable random variables can be explained with some latent variable model under the context of social and behavioral research. Denote a random variable is a column vector for N individuals. Assume there are J outcomes $\mathbf{Y} = (Y_j)_{1\times J}$, P predictors $\mathbf{X} = (X_p)_{1\times P}$, and K latent variables or factors $\mathbf{F} = (F_k)_{1\times K}$ in the model, where Y_j, X_p, and F_k are each an $N \times 1$ random variable with subscripts $j = 1, \ldots, J$, $p = 1, \ldots, P$, and $k = 1, \ldots, K$. Typically, the outcomes are scale items or indicators with the factors as the underlying latent variables, and the predictors are observed demographic or background variables. Since the intercepts are of no interest in this research, all observed variables are assumed to be mean centered to simplify the formulation.

Under the PCLVM framework, both structural and measurement components are integrated into one model with the following equations:

$$\mathbf{Y} = \mathbf{F}\mathbf{A}' + \mathbf{X}\mathbf{H}' + \mathbf{E}, \tag{1}$$

and

$$\mathbf{F} = \mathbf{X}\mathbf{B}' + \mathbf{D}, \tag{2}$$

where $\mathbf{E} = (E_j)_{1\times J}$ and $\mathbf{D} = (D_k)_{1\times K}$ are the J measurement errors and K factor disturbances, respectively, and are assumed to be normally distributed with $\mathbf{E} \sim MVN(0, \mathbf{V})$ and $\mathbf{D} \sim MVN(0, \mathbf{U})$. Model parameters are included in the following matrices:

- $\mathbf{A} = (a_{jk})$ is a $J \times K$ loading matrix;
- $\mathbf{H} = (h_{jp})$ is a $J \times P$ differential coefficient matrix;

- $\mathbf{B} = (b_{kp})$ is a $K \times P$ structural coefficient matrix;
- $\mathbf{V} = (v_{jj'})$ is a $J \times J$ error covariance matrix;
- $\mathbf{U} = (u_{kk'})$ is a $K \times K$ disturbance covariance matrix.

All elements in matrices $\mathbf{A}, \mathbf{H}, \mathbf{B}$ and off-diagonal elements in $\mathbf{V}$ can be fixed (e.g., 0 or 1), freely estimated, or regularized when specification is unclear. On the measurement level, factors are operationally defined by nonzero fixed or freely estimated loadings in $\mathbf{A}$. All elements in $\mathbf{U}$ can be nonzero to allow for correlated disturbances. For $\mathbf{V}$, when off-diagonal elements are assumed to be nonzero and regularized, it is equivalent to assume LD with correlated measure errors.

The indeterminacy of the latent scale can be fixed with two typical options. First, one loading per factor is fixed as one, and the corresponding item is called the reference indicator. It sets the latent scale to the item and is usually referred to as the original solution. Second, factor variance is fixed as one, which is referred to as the standardized solution when the observed variables are also standardized. This option is more complex technically since we need to standardize the factors as dependent variables. But it is also more useful practically, especially when we are unclear which items can be used as reference, and will be the default option in this research. The two options are mathematically equivalent due to scale invariance.

In its most general form, all parameters in $\mathbf{A}, \mathbf{H}, \mathbf{B}$ and off-diagonal elements in $\mathbf{V}$ can be unspecified and regularized separately or jointly with the L_1 norm or Lasso by penalizing the log-likelihood (LLK) function as:

$$LLK + \lambda_A \sum_{a_{jk} \in \mathbf{A}^*} |a_{jk}| + \lambda_B \sum_{b_{kp} \in \mathbf{B}^*} |b_{kp}| + \lambda_H \sum_{h_{jp} \in \mathbf{H}^*} |h_{jp}| + \lambda_V \sum_{v_{jj'} \in \mathbf{V}^*} |v_{jj'}| \tag{3}$$

where $LLK = \frac{2}{N} \log|\mathbf{V}| - \frac{1}{2} tr(\mathbf{SV})$ with $\mathbf{S} = (\mathbf{Y} - \mathbf{FA}' - \mathbf{XH}')'(\mathbf{Y} - \mathbf{FA}' - \mathbf{XH}')$, and $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{H}^*$, and $\mathbf{V}^*$ are the subsets of the vectorized matrices $vec(\mathbf{A})$, $vec(\mathbf{B})$, $vec(\mathbf{H})$, and $vec(\mathbf{V})$, respectively, that are unspecified and require Lasso regularization, with λ_A, λ_B, λ_H, λ_V as the corresponding shrinkage parameters. Lasso parameters $\mathbf{A}^*$, $\mathbf{B}^*$, $\mathbf{H}^*$, and $\mathbf{V}^*$ can be selected through different design matrices.

The general formulation is visually depicted in Figure 1, showcasing a model encompassing six indicators, two factors, and two predictors. In this context, the elements within $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{H}$, along with the off-diagonal components within $\mathbf{V}$, represent unspecified Lasso parameters. These parameters are indicated by dashed lines in the figure. This exploratory setting requires only minimal prior knowledge for specification. As more information becomes available, these Lasso parameters can be regularly specified and estimated as free parameters or fixed, as shown by solid lines in the figure. Notably, irrelevant parameters are set to zero and omitted from the model. With full knowledge, the model becomes strictly confirmatory, and all parameters are either specified or fixed. Meanwhile, the framework reduces to the PCFA models when no predictors are included, indicating that predictors are optional. However, by

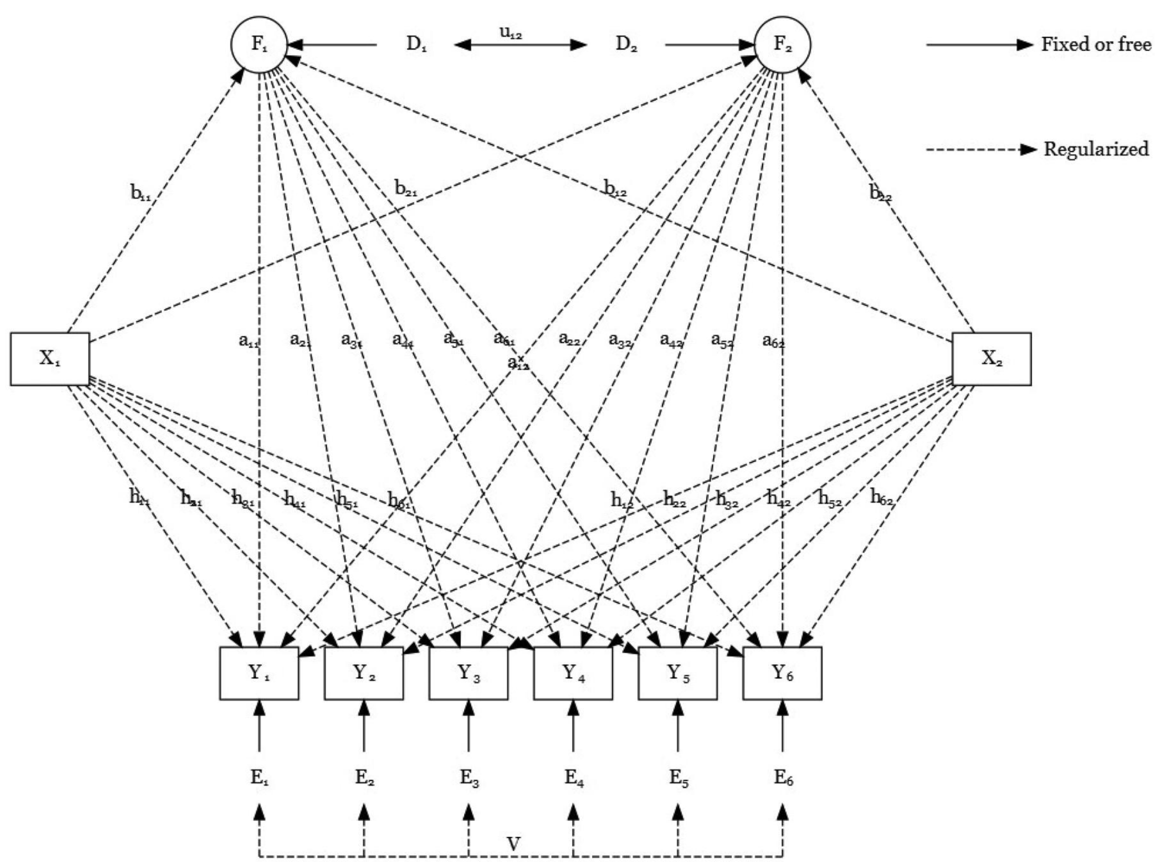


Figure 1. General form.

incorporating predictors, it is possible to regularize two coefficient matrices, thereby enhancing the flexibility and scalability of the framework, as discussed below.

The general formulation, while flexible, presents challenges in terms of identification and estimation without the imposition of additional constraints. The identification of specific designs relies on the application of various combinations of regularizations to the parametric matrices. To estimate these designs effectively, one must have access to complete conditional distributions, which need to be derived analytically and validated empirically. In the following, we concentrate on typical cases that have been successfully identified and are practically valuable as research designs.

3. Specific Design

3.1. Case One

This case exhibits an exploratory tendency, with minimal substantive knowledge available to specify the loading matrix **A** and structural coefficient matrix **B**. **A** can be mostly unspecified and regularized, and **B** can be fully unspecified and regularized while the differential coefficient matrix **H** doesn't exist. Specifically, only minimum knowledge is needed to operationally define each factor (e.g., a few loadings per factor). In practice, it is more likely that each item is designed to measure specific factor and can be specified accordingly. The model remains meaningful for probing cross-loadings and structural coefficients simultaneously, although most of which can be minor. This approach aligns with the concept of more flexible representation in SEM (Muthén & Asparouhov, 2012), where the strict requirement of many zero cross-loadings and structural coefficients under the confirmatory approach can be relaxed. But we offer more flexibility to address uncertainty in both the loading and structural coefficient matrices. This method is particularly useful in large-scale settings with numerous items, factors, or predictors. Note that when the predictors are categorical, it is equivalent to conducting multiple group comparisons and significant structural coefficients imply group difference. In the absence of predictors, the model reduces to the E-step in PCFA.

The technical challenge lies in regularizing two parametric matrices at the measurement and structural levels: **A** can be partially regularized and mixed with Lasso, free and fixed parameters while **B** is typically entirely regularized with Lasso parameters in the context of correlated disturbances (Figure 2).

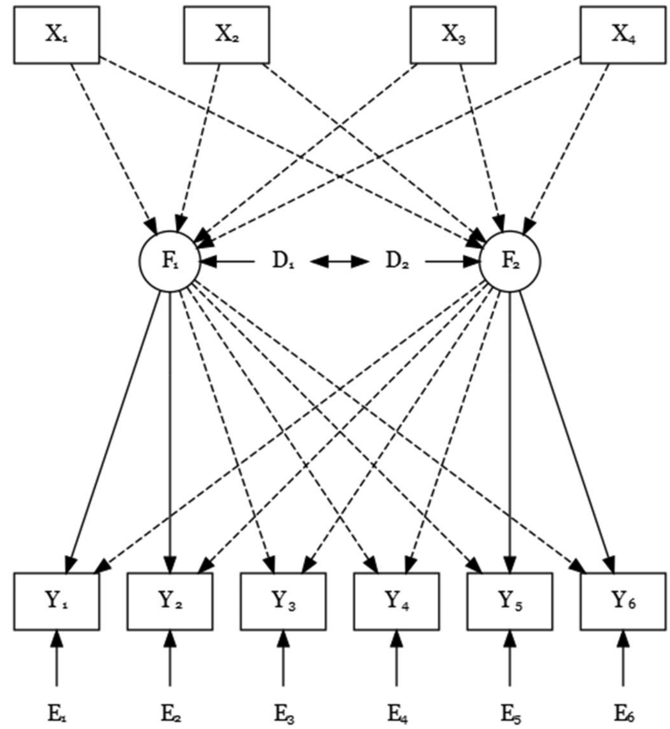


Figure 2. Case one.

3.2. Case Two

This case is confirmatory-inclined with a partially specified loading matrix. All elements in matrices **B** and off-diagonal elements in **V** can be unspecified and regularized, which means that local dependence can be accommodated. Here sufficient knowledge should be available to partially specify **A** with the minimum condition of one loading per item. This case is useful when there are uncertainties in both the structural coefficient and loading matrices, with concern of local dependence. In practice, **B** can be partially specified, resulting in less burden of regularization. When **B** is fully specified or does not exist, the case reduces to the C-step in PCFA in essence (Figure 3).

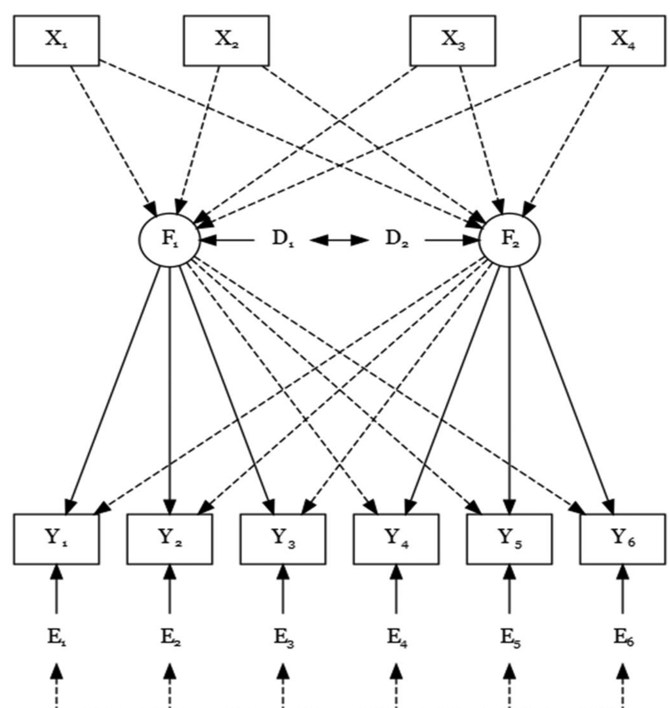


Figure 3. Case two.

3.3. Case Three

This case can be considered a variation of the first case, also exhibiting an exploratory tendency. In this scenario, predictors serve as causal indicators that define the factors and are necessary rather than optional. As a result, each factor is defined externally by the predictors, rather than internally by the items. With the assistance of causal indicators, we can explore the pattern of the loading matrix, which can be entirely unspecified and regularized.

This situation arises when there is limited information about the measurement component, and the model is augmented with ancillary covariates. The loading matrix **A** is entirely regularized with Lasso parameters, while parameters in the structural coefficient matrix **B** are specified as free or fixed as zero. It is important to note that only the standardized solution is viable, as there are no items available to serve as reference indicators (Figure 4).

3.4. Case Four

This case represents a scenario where the measurement component including all item residual correlations are entirely exploratory, while the structural component is fully specified. This case is suitable for situations when there is a need to address potential local dependencies under Case Three (Figure 5).

3.5. Case Five

This design can be effectively utilized to explore the potential existence of structural effects, cross-loadings, and differential item functioning (DIF) effects concurrently. It's suitable for exploring DIF and measurement invariance in complex scenarios and can serve as a follow-up to Case 1 for further model investigation. The presence of any significant differential coefficient h_{jp} suggests that the probability of endorsing an item will differ for students with the same factor level but different predictor levels. In other words, there is a DIF effect of predictor p on item j.

The identification of structural effects can provide policy makers with valuable insights into significant differences between student groups. However, a substantial presence of DIF effects can challenge the assumption of measurement invariance, which is paramount for group comparisons (Bauer, 2017). PCLVM offers a more effective way to simultaneously investigate many possible DIF effects in a large-scale system compared to traditional analysis.

In this design, sufficient knowledge should be available to partially specify matrix **A**, with a minimum condition of one loading per item. It provides a flexible range, from a more exploratory approach (specifying few loadings per factor) to a more confirmatory approach (specifying one loading per item). In practice, it's common to specify most major loadings per factor, making the case more confirmatory. The model remains flexible and useful by accommodating unclear cross-loadings and structural coefficients while simultaneously exploring possible DIF effects.

Technically, this case involves the regularization of three parametric matrices: **A** and **H** at the measurement level and **B** at the structural level. Specifically, **A** can be partially regularized and mixed with Lasso, free and fixed parameters while **H** is typically regularized entirely. Meanwhile, **B** can be entirely regularized with Lasso parameters and correlated disturbances. Managing three matrices for regularization at two levels makes the algorithm complex and challenging, requiring univariate and multivariate Bayesian Lassos involving correlated residuals, which are not straightforward

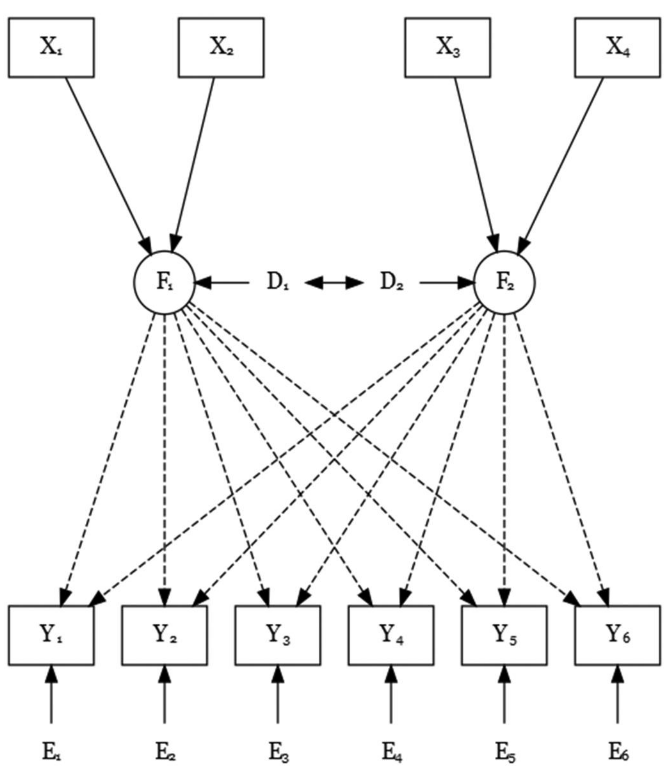


Figure 4. Case Three.

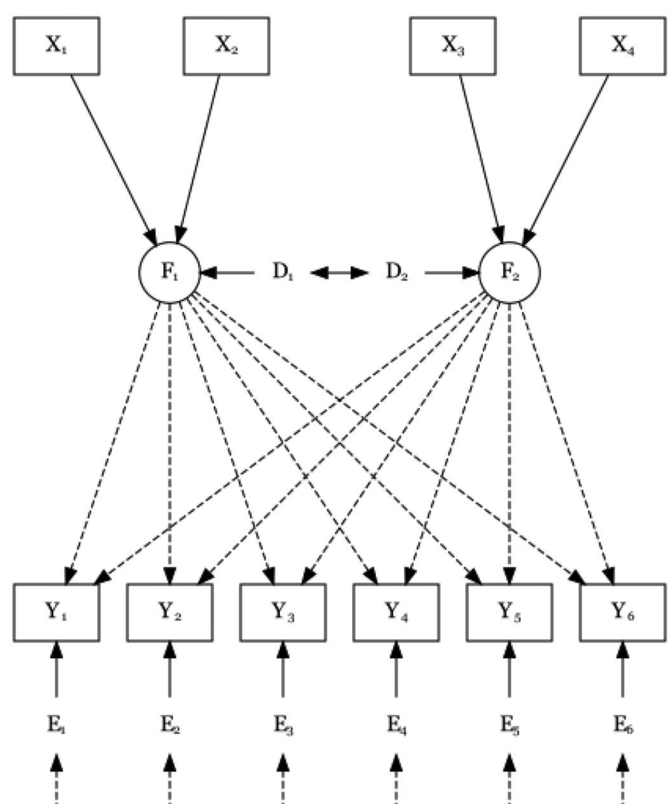


Figure 5. Case Four.

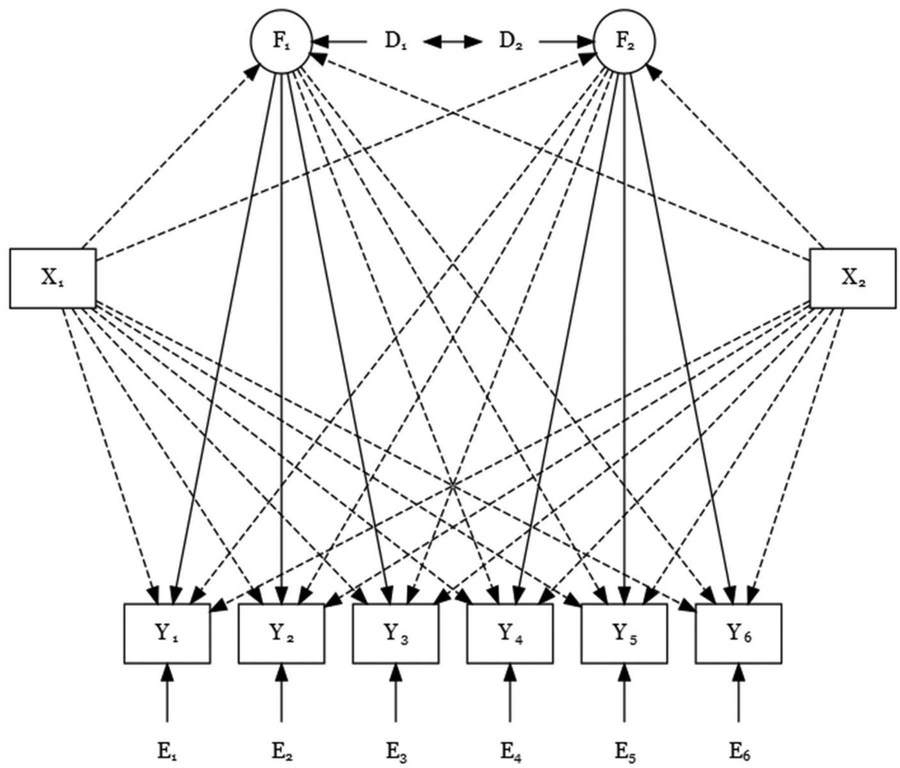


Figure 6. Case five.

to derive analytically. Alternatively, **B** can be fully specified with zero-fixed and free parameters based on results from Case 1. This approach slightly simplifies the algorithm with only two matrices to regularize (Figure 6).

3.6. Case Six

This design is effective for investigating both structural and differential item functioning (DIF) effects, along with correlated residuals at both the structural and measurement levels. It can be seen as a variation of the fifth case, with the inclusion of local dependence. As shown in the figure, the loading matrix **A** should be partially specified with a minimum condition of one loading per item, while structural coefficients in **B**, differential coefficients in **H**, and off-diagonal elements in **V** can be unspecified and regularized. This case is particularly valuable when both DIF effects and local dependence are areas of concern.

Technically, this case entails the regularization of several parametric matrices: **A** and **H** at the measurement level and **B** at the structural level. The algorithm for this case is similarly complex and challenging and can be combined with univariate and multivariate Bayesian Lassos, taking into account correlated residuals at both levels. Specifications for the loading matrix can be based on analyses from Cases 1 or 3. As in Case 5, **B** can be fully specified with zero-fixed and free parameters, which simplifies the algorithm by reducing it to just three matrices that need regularization (Figure 7).

3.7. Case Seven

This design serves as an extension of traditional bifactor models, with regularization applied to a partially specified loading matrix and a fully unspecified structural coefficient matrix in the context of orthogonal factors. The general factor F_1 represents a fundamental construct and is measured by all items. The special factors, F_2 and F_3, represents factors resulting from special characteristics of the model such as method, format, or testlet effects. While each special factor is typically defined by a subset of the items by design, regularization can help determine whether it is unintentionally measured by other items, providing insights for scale development. This case is also valuable for evaluating the impact of different predictors on general or special factors. In the absence of predictors, the model reduces to the PCFA for the bifactor setting, which can be used to assess the effect sizes of the special factors.

The technical algorithm is relatively straightforward, as the special factors are usually orthogonal to each other and the general factor, resulting in uncorrelated disturbances. However, this case can be extended to a more complex scenario where multiple general factors exist, and their disturbances are correlated with each other (though not with those of the special factors). This extension allows for the exploration of more intricate relationships in the model (Figure 8).

3.8. Case Eight

This design can be seen as a variation of the seventh case, with accommodation of local dependence but exclusion of

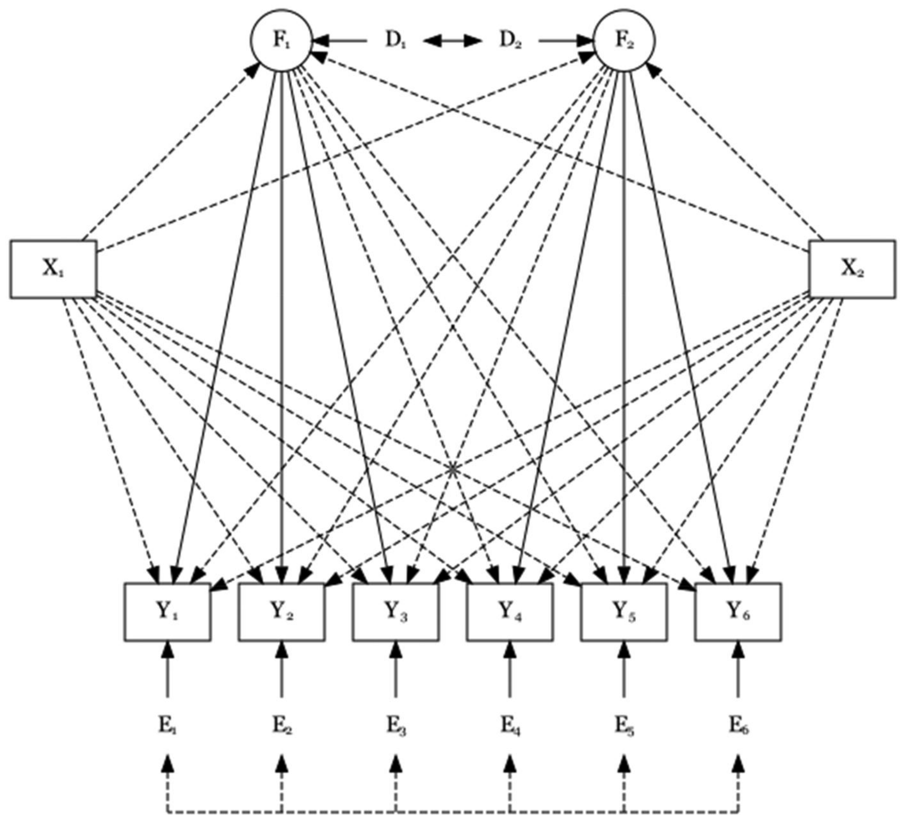


Figure 7. Case six.

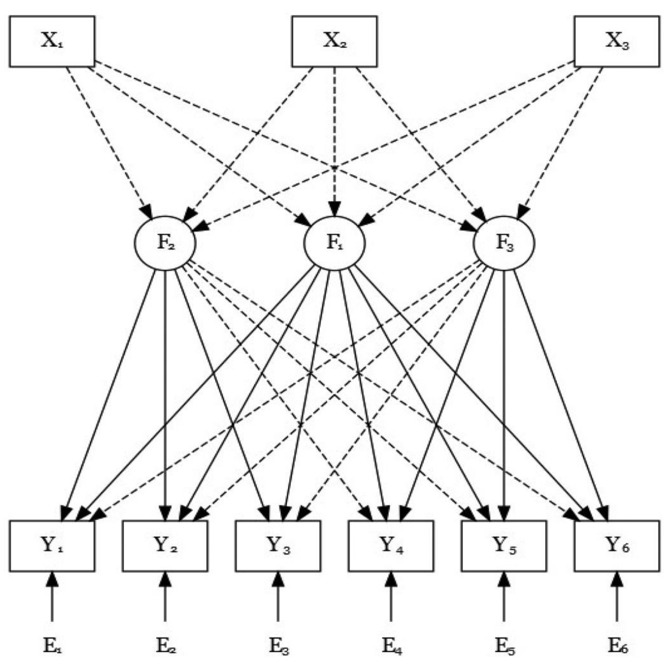


Figure 8. Case seven.

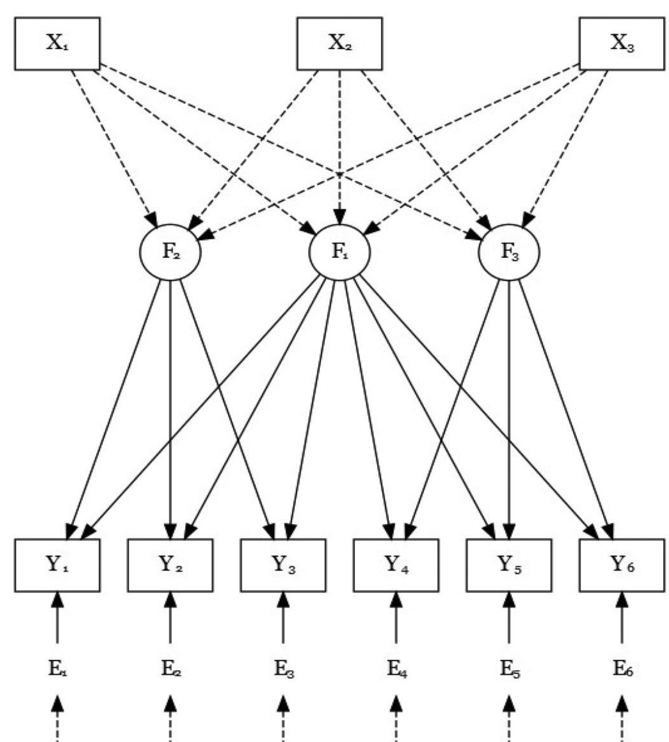


Figure 9. Case eight.

loading regularization. Specifically, the loading matrix **A** should be fully specified, while structural coefficients in **B** and off-diagonal elements in **V** can be fully unspecified and regularized. This case is particularly valuable when researchers are concerned about both structural effects and local dependence under a bifactor setting. Like the seventh case, this design can be extended to handle more complex scenarios where multiple general factors exist, and their disturbances are correlated with each other (but not with those of the special factors). This extension offers greater flexibility in exploring relationships within the model (Figure 9).

3.9. Bayesian Modeling and Estimation

The key to understanding the Bayesian Lasso is to turn the penalty part in Equation (3) into priors, with the shrinkage parameter as hyper-prior. Moreover, when appropriate priors are selected to maintain conjugacy, full conditional distributions are available for effective Gibbs sampler. The Bayesian model in hierarchical form can be developed as follows.

For item j, denote the row vectors of the loading matrix $\mathbf{A}$ and differential coefficient matrix $\mathbf{H}$ as $\mathbf{a}_j$ and $\mathbf{h}_j$, respectively, for $j = 1, \ldots, J$. Denote $\mathbf{a}_j^*$ and $\mathbf{a}_j^{**}$ as the subsets of Lasso and free parameters for $\mathbf{a}_j$, respectively, with K_j^* and K_j^{**} as corresponding lengths. Note that $K_j^* + K_j^{**} \le K$ since there can be fixed parameters in $\mathbf{a}_j$. Equation (1) can be reformulated as:

$$\mathbf{Y} \sim MVN(\mathbf{FA}' + \mathbf{XH}', \mathbf{V}), \tag{4}$$

with

$$\mathbf{a}_j^{**} \sim N\left(\boldsymbol{\mu}_{Aj}, \boldsymbol{\Sigma}_{Aj}\right), \tag{5}$$

where $\boldsymbol{\mu}_{Aj}$ and $\boldsymbol{\Sigma}_{Aj}$ are the hyper-priors. Meanwhile,

$$\mathbf{a}_j^{*} \sim N\left(0, \mathbf{T}_{Aj}\right), \tag{6}$$

with

$$\mathbf{T}_{Aj} = \text{diag}\left(\tau_{Aj1}^2, \ldots, \tau_{AjK_j^*}^2\right), \tag{7}$$

$$\tau_{Aj1}^2, \ldots, \tau_{AjK_j^*}^2 \sim \prod_{k=1}^{K_j^*} \frac{\lambda_A^2}{2} \exp\left(-\lambda_A^2 \tau_{Ajk}^2 / 2\right) d\tau_{Ajk}^2, \tag{8}$$

and

$$\lambda_A^2 \sim Gamma(\alpha_A, \beta_A), \tag{9}$$

where typically $\alpha_A = 1$ and $\beta_A > 0$ (but small) are the shape and rate parameters, respectively (same below).

Different from $\mathbf{a}_j$, the differential coefficient vector $\mathbf{h}_j$ is usually entirely regularized:

$$\mathbf{h}_j \sim N\left(0, \mathbf{T}_{Hj}\right), \tag{10}$$

with

$$\mathbf{T}_{Hj} = \text{diag}\left(\tau_{Hj1}^2, \ldots, \tau_{HjP}^2\right), \tag{11}$$

$$\tau_{Hj1}^2, \ldots, \tau_{HjP}^2 \sim \prod_{p=1}^{P} \frac{\lambda_H^2}{2} \exp\left(-\frac{\lambda_H^2 \tau_{Hjp}^2}{2}\right) d\tau_{Hjp}^2 \tag{12}$$

and

$$\lambda_H^2 \sim Gamma(\alpha_H, \beta_H). \tag{13}$$

In LD cases (i.e., Case 2, 4, 6, and 8), the residual covariance matrix $\mathbf{V}$ is non-diagonal and Bayesian covariance Lasso will be used to regularize off-diagonal elements with double exponential priors:

$$\begin{aligned} p\left(v_{jj'} | \tau_{Vjj'}\right)_{j<j'} &\propto \tau_{Vjj'}^{-\frac{1}{2}} \exp\left(-\frac{v_{jj'}^2}{2\tau_{Vjj'}}\right), \\ p\left(v_{jj} | \lambda_V\right) &\propto \frac{\lambda_V}{2} \exp\left(-\frac{\lambda_V v_{jj}}{2}\right), \\ p\left(\tau_{Vjj'} | \lambda_V\right)_{j<j'} &\propto \frac{\lambda_V^2}{2} \exp\left(-\frac{\lambda_V^2 \tau_{Vjj'}}{2}\right), \end{aligned} \tag{14}$$

with $\lambda_V \sim Gamma(\alpha_V, \beta_V)$. In no-LD cases (i.e., Case 1, 3, 5, and 7), $\mathbf{V}$ is diagonal:

$$v_{jj}^{-1} \sim Gamma(\alpha_V, \beta_V). \tag{15}$$

The above equations represent the hierarchical formulation of the univariate Bayesian Lasso (e.g., Kyung et al., 2010; Park & Casella, 2008), where there is individual τ^2 for each Lasso parameter. After integrating out τ^2, each Lasso parameter exhibits the desired conditional Laplace (i.e., double-exponential) prior. However, such hierarchy is unsuitable for multivariate cases with correlated factorial residuals, as found in the structural model. Instead, we will formulate the structural model with a multivariate version of the Bayesian Lasso (Culpepper & Park, 2017), as follows:

$$\mathbf{F} \sim N(\mathbf{XB}', \mathbf{U}), \tag{16}$$

Upon vectorizing $\mathbf{B}$, it can be fully regularized conditional on $\mathbf{U}$:

$$vec(\mathbf{B}) | \mathbf{U} \sim N\left(0, \mathbf{U} \otimes \mathbf{T}_B\right), \tag{17}$$

where

$$\mathbf{T}_B = \text{diag}\left(\tau_{B1}^2, \ldots, \tau_{BP}^2\right), \tag{18}$$

$$\tau_{Bp}^2 | \lambda_B^2 \sim Gamma\left[(K+1)/2, \lambda_B^2/2\right], \tag{19}$$

and

$$\lambda_B^2 \sim Gamma(\alpha_B, \beta_B). \tag{20}$$

Here, $\otimes$ denotes the Kronecker product, and α_B and β_B are hyper-priors. This implies that each structural vector $\mathbf{b}_p = \left(b_{1p}, \ldots, b_{Kp}\right)$ will share the same τ_{Bp}^2 for $p = 1, \ldots, P$. This shared structure is necessary for performing Lasso regularization when $\mathbf{U}$ is non-diagonal as demonstrated in Appendix A. After integrating out τ_{Bp}^2, the conditional prior of $\mathbf{b}_p$ is proportional to a multivariate generalized asymmetric Laplace distribution (Kozubowski et al., 2013).

For Cases 2 and 3, we only need to update the free parameters in $vec(\mathbf{B})$, as $\mathbf{B}^{**}$:

$$\mathbf{B}^{**} | \mathbf{U} \sim N\left(\boldsymbol{\mu}_B, \mathbf{U} \otimes \boldsymbol{\Sigma}_B\right), \tag{21}$$

where $\boldsymbol{\mu}_B$ and $\boldsymbol{\Sigma}_B$ are the hyper-priors.

$$\mathbf{U}^{-1} \sim Wishart(\delta_U, \boldsymbol{\Sigma}_U), \tag{22}$$

In Case 7, the disturbances are uncorrelated, which means there are only diagonal elements in $\mathbf{U}$:

$$u_{kk}^{-1} \sim Gamma(\alpha_U, \beta_U). \tag{23}$$

With the above formulation, the full conditional distributions of all parameters can be obtained as shown in Appendix A, which can be used for MCMC estimation. An effective Gibbs sampler can be implemented by iteratively sampling from the conditional distributions. To obtain the standardized solution, one can keep standardizing $\mathbf{F}$ during each iteration, noting that $\mathbf{Y}$ and $\mathbf{X}$ should be standardized before sampling. Alternatively, one can also obtain the standardized solution post-hoc by rescaling the parameters with the variances of $\mathbf{F}, \mathbf{Y}, \mathbf{X}$ after sampling.

Bayesian Lasso cannot give exactly zero estimates for the Lasso parameters, which is needed to achieve model selection under the frequentist approach. One solution is to create Bayesian credible intervals to assess statistical significance for the parameters. The highest posterior density (HPD) intervals based on the MCMC iterations after the burn-in period is adopted. Posterior predictive (PP) p value can be used as a complementary statistic, considering Bayesian lasso is already a model selection technique. Chain convergence can be determined using the estimated potential scale reduction (EPSR), with 1.1 as the criterion. Gibbs sampler (Geman & Geman, 1984) is adopted as the estimation algorithm. More technical details of implementation can be found in Appendix A.

4. Simulation Studies

Simulation studies were conducted to evaluate the performance of the PCLVM and comparisons with ESEM on Cases 1 to 4. For data generation with or without LD, we set two population models with $P=9$, $J=18$, $K=3$, and $N=1{,}000$. True model parameters can be found in Figures B1 and B2 in Appendix B, which is a symmetric design with each factor measured by six major loadings and regressed on three predictors. Meanwhile, LD terms were designed to be equally distributed within and between factors. For each simulation case, 100 datasets were simulated and analyzed.

4.1. Study One: Performance of the PCLVM

PCLVM models for Case 1 to 4 were used to recover the two population models. In Case 1 and 2, respectively, the first three and all major loadings for each factor were specified as free parameters while all other parameters in **A** and **B** were unspecified as Lasso parameters. In Case 3 and 4, the three nonzero coefficients in **B** were specified as free parameters with other coefficients fixed as zero. All parameters in **A** were unspecified as Lasso parameters. Note that, for LD cases (i.e., Cases 2 and 4), all off-diagonal elements in **V** were unspecified and regularized with Lasso.

In the MCMC estimation, EPSRs were found to be no more than 1.1 within 10,000 iterations which were used as burn-in. Another 10,000 draws were used to obtain the conditional distributions of parameter estimates. With the distributions, performance assessment was conducted based on the bias of the median estimates (BIAS) between the estimates and true values, the related root mean square error (RMSE), the mean of the standard error estimates (SE), and the percentage of significant estimates at α =.05 based on the HPD interval (SIG%).

Simulation results were shown in Table 1, which were satisfactory in general (e.g., in terms of BIAS, RMSE, SE, or SIG%). When data were generated with LD, Cases 2 and 4 with regularized **V** performed slightly better than Cases 1 and 3, which ignored the LD structure. While there were some Type I errors (i.e., the SIG% of zero parameters) in all cases, the magnitudes of the inflated estimates were generally small and well below .1 (either the loading or structural coefficient).

Table 1. PCLVM parameter estimates in Cases 1 to 4.

		Data Generation w/o LD				Data Generation with LD			
Par	TRUE	BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
					Case 1				
a	0.7	0.005	0.021	0.032	1.000	−0.002	0.094	0.033	1.000
a_0	0	−0.007	0.027	0.051	0.000	−0.015	0.045	0.059	0.023
b	0.3	0.009	0.029	0.020	1.000	0.031	0.041	0.020	1.000
b_0	0	0.001	0.030	0.027	0.101	−0.001	0.030	0.027	0.097
u	0.3	−0.022	0.027	0.062	1.000	−0.049	0.050	0.071	1.000
					Case 2				
a	0.7	0.011	0.026	0.037	1.000	0.008	0.029	0.038	1.000
a_0	0	−0.011	0.028	0.053	0.000	−0.019	0.033	0.052	0.003
b	0.3	0.007	0.029	0.020	1.000	0.007	0.029	0.020	1.000
b_0	0	0.003	0.029	0.027	0.099	0.006	0.030	0.027	0.104
u	0.3	−0.032	0.034	0.064	1.000	−0.040	0.042	0.067	1.000
v	0.3					−0.024	0.038	0.035	1.000
					Case 3				
a	0.7	0.005	0.023	0.029	1.000	0.020	0.037	0.035	1.000
a_0	0	−0.006	0.035	0.038	0.030	−0.010	0.036	0.041	0.031
b	0.3	0.000	0.030	0.019	1.000	−0.007	0.036	0.020	1.000
u	0.3	−0.015	0.035	0.048	1.000	−0.029	0.045	0.050	0.963
					Case 4				
a	0.7	−0.001	0.095	0.028	1.000	0.010	0.032	0.036	1.000
a_0	0	−0.017	0.050	0.041	0.115	−0.013	0.037	0.041	0.030
b	0.3	0.021	0.035	0.018	1.000	−0.003	0.031	0.019	1.000
u	0.3	−0.049	0.055	0.054	1.000	−0.020	0.040	0.052	0.990
v	0.3					−0.022	0.040	0.035	1.000

Note. 'a' averaged across all major loadings; 'a_0' averaged across all zero loadings; 'b' averaged across all non-zero structural coefficients; 'b_0' averaged across all zero structural coefficients; 'u' averaged across all factorial residual covariance; 'v' averaged across all LD terms.

Table 2. ESEM parameter estimates in Cases 1 to 4.

		Data Generation w/o LD				Data Generation with LD			
Par	TRUE	BIAS	RMSE	SE	SIG%	BIAS	RMSE	SE	SIG%
					Case 1				
a	0.7	0.019	0.026	0.020	1.000	0.010	0.093	0.023	1.000
a_0	0	−0.054	0.059	0.025	0.593	−0.061	0.072	0.026	0.604
b	0.3	−0.011	0.030	0.029	1.000	0.013	0.030	0.029	1.000
b_0	0	0.022	0.037	0.030	0.109	0.016	0.034	0.030	0.080
u	0.3	−0.057	0.057	0.024	1.000	−0.104	0.104	0.022	1.000
					Case 2				
a	0.7	0.035	0.087	NA	NA	0.029	0.105	NA	NA
a_0	0	−0.113	0.235	NA	NA	−0.119	0.254	NA	NA
b	0.3	0.015	0.044	NA	NA	0.028	0.052	NA	NA
b_0	0	0.040	0.104	NA	NA	0.039	0.107	NA	NA
u	0.3	−0.298	0.298	NA	NA	−0.298	0.298	NA	NA
v	0.3					0.314	0.335	NA	NA

Note. NA: (SE) not available.

4.2. Study Two: Comparison with ESEM

Surprisingly, we found that ESEM with target rotation implemented in Mplus (Muthén & Muthén, 2017) can only be applied to Cases 1 to 2 in the research designs. Moreover, while all elements in the residual matrix **V** can be estimated in Case 2, there is no regularization per se. Parameter estimates were shown in Table 2. Most estimates for both Cases were slightly worse than, but comparable to, PCLVM. The issues in Case 1 were the large Type I error rates for the zero loadings. More importantly, no SE can be estimated in either Case 2, suggesting that the point estimates are unreliable or not trustworthy.

5. Real-Life Data Analysis

PCLVM is particularly beneficial in settings with numerous observable and latent variables, such as large-scale assessments featuring many background variables and scale items. The Trends in International Mathematics and Science Study (TIMSS) is an example of such a setting. Appendix C presents scale items measuring three factors related to 8th grade mathematics learning in TIMSS 2019 (IEA, 2018): intrinsic motivation (F_1), self-efficacy (F_2), and extrinsic motivation (F_3). Although each item is clearly designed to measure specific factor with major loading, unexpected cross-loadings are likely and merit further investigation. Additionally, Appendix C displays a range of demographic variables that could impact these factors. Three PCLVM models, Cases 1, 2, and 6 can be adopted to fit the data. In Case 1, we only specify three major loadings per factor, which allows for the simultaneous identification of loading uncertainty, significant cross-loadings, and structural coefficients. Building on the above basic analysis, we can further investigate the local dependence of the residuals between items with Case 2 and differential coefficients of the predictors on the items for potential DIF effects with Case 6. For ESEM, only Case 1 is applicable. Responses from 2,073 English students, after removing missing responses, are used for illustration.

Table 3 shows that all PCLVM-suggested models fit the data better than ESEM. It is worth noting that, while Case 6 presents the best model, Case 2 is also acceptable and probably a better choice if the DIF effect is not of major concern.

Loading and structural coefficient estimates for Case 6 can be found in Table 4. Most major loadings are estimated as expected, along with some significant cross-loadings which are small either in an absolute sense or in comparison to the major loadings. The only exception is item Y_{16}, which may require revision. Except for four predictors, the magnitudes of significant structural coefficients are smaller than .1 and neglectable. Of special interests are X_{13}, X_{17}, X_{19}, and X_{23}, and the implications are quite straightforward except for X_{13}. The negative effect of how far in education one is expected to go (X_{13}) on self-efficacy and extrinsic motivation is counterintuitive. Ultimately, the results highlight the importance of independent work (X_{19}) in mathematics education.

Significant LD terms can be found in Table 5, the magnitudes of most are smaller than .1 and can be ignored. It is interesting to see that all LD terms larger than .1 occur within the same factors. The significant DIF effects can be found in Appendix D, the magnitudes of all are smaller than .1. It suggests that measurement invariance is not a major concern for this set of predictors.

6. Discussion

The LVM framework is a powerful tool for researchers, providing different approaches to address major research processes. Within the LVM framework, exploratory and confirmatory methods represent opposite ends of a continuum, and researchers often need support to navigate this continuum effectively. The challenge lies in finding an approach that can comprehensively cover the entire spectrum. Recently, regularization methods have been developed to provide researchers with greater flexibility in integrating varying levels of substantive knowledge and accommodating a broader range of the confirmatory-exploratory continuum. This research builds upon these developments by introducing predictors and regularizations to two additional parameter matrices: structural and differential coefficients. The outcome is a comprehensive framework, where researchers can apply different regularizations to these parameter matrices individually or collectively, and in full or in part.

With PCLVM, applied researchers can design a variety of research studies for different purposes, depending on the combinations of different regularizations. The attractiveness of the proposed framework was demonstrated through a variety of typical cases that can be readily estimated and widely encountered in practice. The diversity of research designs enhances the versatility of this approach, making it a valuable tool for addressing a wide array of research questions and hypotheses in different contexts. For methodologists, PCLVM provides more flexibility when there's uncertainty related to any of the four parametric matrices. One can also explore different combinations of regularizations for their own purpose under the framework. Through both simulation studies and real-data analysis, PCLVM demonstrates clear advantages against ESEM. Specifically, ESEM can only be implemented in two of the eight cases, with meaningful results in one case. PCLVM also

Table 3. Model fit of different models.

Model	RMSE	CFI	TLI	SRMR	BIC	LLK	DF
PCLVM-Case 1	0.059	0.803	0.787	0.048	116449.8	−57629.23	1119
PCLVM-Case 2	0.04	0.906	0.893	0.048	119206	−58816.44	1069
PCLVM-Case 6	0.035	0.931	0.922	0.032	105895.2	−52237.38	974
ESEM-Case 1	0.064	0.761	0.741	0.057	153097.2	−75868.94	1202

Table 4. Significant loading and structural coefficient estimates.

Item	F_1	F_2	F_3	Var	F_1	F_2	F_3
Y_1	0.774			X_1			−0.060
Y_2	0.564	0.116	0.081	X_2			
Y_3	0.720			X_3	−0.049		
Y_4	0.625		0.142	X_4			
Y_5	0.820			X_5			
Y_6	0.717			X_6			
Y_7	0.736			X_7			
Y_8	0.820			X_8			0.079
Y_9	0.748	0.131		X_9			−0.027
Y_{10}		0.739		X_{10}			−0.050
Y_{11}		0.755		X_{11}			0.049
Y_{12}		0.799		X_{12}		0.040	−0.040
Y_{13}		0.734		X_{13}	**−0.111**	**−0.152**	**−0.218**
Y_{14}		0.565		X_{14}	−0.084	−0.044	−0.041
Y_{15}		0.696		X_{15}	−0.021	−0.021	0.017
Y_{16}	0.149	0.275		X_{16}	−0.043	−0.052	−0.038
Y_{17}		0.793		X_{17}	**−0.177**	**−0.120**	−0.073
Y_{18}		0.693		X_{18}	−0.078	−0.079	−0.028
Y_{19}			0.603	X_{19}	**0.227**	**0.370**	**0.156**
Y_{20}			0.611	X_{20}			
Y_{21}	−0.130		0.760	X_{21}	−0.065	0.032	−0.061
Y_{22}			0.795	X_{22}		−0.031	
Y_{23}	0.203	0.147	0.453	X_{23}	**−0.115**	**−0.205**	**−0.144**
Y_{24}			0.798	X_{24}	−0.036	−0.014	
Y_{25}			0.835	X_{25}	−0.072		−0.067
Y_{26}	−0.128		0.538	X_{26}	0.041	0.096	−0.027
Y_{27}			0.728				

Note. Only significant estimates are presented; Major loadings in measurement part are underscored; Absolute values above .1 in structure part are highlighted.

Table 5. Significant LD terms.

	Y_1	Y_2	Y_3	Y_4	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{19}	Y_{20}	Y_{21}	Y_{22}	Y_{23}	Y_{26}
Y_3		.099																				
Y_4	.038																					
Y_5	.067																					
Y_7					**.115**																	
Y_9							.071															
Y_{11}									−.065													
Y_{12}				−.043						**.113**												
Y_{13}		−.040									−.042											
Y_{14}									−.056	.069		−.044										
Y_{15}		−.038			.034	.064				−.047	−.034	.076	−.046									
Y_{16}				.045			.062		.073	−.069	−.066		−.062									
Y_{17}									−.054	.096	.093	−.042	**.109**	−.062	−.074							
Y_{18}			.053						−.050				.089	−.047	−.067	.099						
Y_{19}		.049		.061							−.040											
Y_{20}				.062							−.036						**.171**					
Y_{21}																	−.099					
Y_{22}																	−.062	−.050	**.105**			
Y_{23}					.042			.043												**.103**		
Y_{24}																	.091		−.092	−.052		
Y_{26}																	−.051				−.042	
Y_{27}																					−.075	**.154**

Note. Absolute values above .1 are highlighted. Items with no significant correlated residuals are not displayed.

outperforms ESEM in all cases with either simulated or real-life data. However, existing population models for data generation are relatively simple, and the performance of PCLVM under more complex situations remains questionable.

In terms of future directions, the framework is scalable and can be readily extended to accommodate categorical responses, missing data, and other psychometric issues, similar to the extension of the PCFA models to generalized PCFA models (Chen, 2021b). Additionally, a significant number of the designs can be further extended by incorporating the mediation or moderation analysis. These are two important techniques that can enhance interpretability and causality in statistical inference (Hayes, 2017). These techniques can also be used in combination to investigate sophisticated causal pathways and identify nuanced patterns under complex settings. But mediation and moderation analyses with latent variables are challenging by themselves, less to mention the incorporation of regularizations at the structural and measurement levels. Moreover, one should exercise caution regarding identification issues when the unspecified pattern is complex or there is an abundance of unspecified parameters.

Finally, the Bayesian implementation based on MCMC can be time consuming and less efficient, especially for large-scale models with numerous items, factors, and/or predictors. For instance, models with sample sizes like those in real-life data analysis will take more than one minutes for 10,000 iterations, even with the effective Gibbs sampler, on a PC with an Intel Core i7 CPU. Investigation of more efficient algorithms for Bayesian inference such as variational approximation (e.g., Neal & Hinton, 1998; Saul & Jordan, 1995) is desirable in the future. Recent research has highlighted the value of mean-field variational methods, introduced by Hinton and Van Camp (1993), under the context of factor analysis. Khan et al. (2010) introduced the variational expectation-maximization algorithm for fitting factor analysis models, and Wang et al. (2020) expanded upon this by integrating the method with a regularization approach. When combined with efficient variational methods, the potential benefits of the PCLVM framework can be fully unlocked, especially to address the perceived opaqueness nature of machine-learning algorithms for large scale models or big data.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a General Research Fund Grant (17603022) from the Hong Kong Research Grants Council.

ORCID

Jinsong Chen http://orcid.org/0000-0002-0157-5469
Yifan Zhang http://orcid.org/0009-0002-7345-2561

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

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

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information Processing Letters*, *24*, 377–380. https://doi.org/10.1016/0020-0190(87)90114-1

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*, 167–174. https://doi.org/10.1080/00031305.1992.10475878

Chen, J. (2020). A partially confirmatory approach to the multidimensional item response theory with the Bayesian Lasso. *Psychometrika*, *85*, 738–774. https://doi.org/10.1007/s11336-020-09724-3

Chen, J. (2021a). A Bayesian regularized approach to exploratory factor analysis in one step. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 518–528. https://doi.org/10.1080/10705511.2020.1854763
Chen, J. (2021b). A generalized partially confirmatory factor analysis framework with mixed Bayesian Lasso methods. *Multivariate Behavioral Research*, *57*, 879–894. https://doi.org/10.1080/00273171.2021.1925520
Chen, J. (2022). Partially confirmatory approach to factor analysis with Bayesian learning: A LAWBL tutorial. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 800–816. https://doi.org/10.1080/10705511.2022.2039660
Chen, J. (2023). Fully and partially exploratory factor analysis with bi-level Bayesian regularization. *Behavior Research Methods*, *55*, 2125–2142. https://doi.org/10.3758/s13428-022-01884-7
Chen, J., Guo, Z., Zhang, L., & Pan, J. (2021). A partially confirmatory approach to scale development with the Bayesian Lasso. *Psychological Methods*, *26*, 210–235. https://doi.org/10.1037/met0000293
Culpepper, S. A., & Park, T. (2017). Bayesian estimation of multivariate latent regression models: Gauss versus Laplace. *Journal of Educational and Behavioral Statistics*, *42*, 591–616. https://doi.org/10.3102/1076998617700598
Feng, X.-N., Wu, H.-T., & Song, X.-Y. (2017). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 341–358. https://doi.org/10.1080/10705511.2016.1257353
Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelharter (Eds.), *Markov Chain Monte Carlo in practice* (pp. 131–144). Chapman & Hall.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Chapman & Hall.
Gelman, A., Meng, X.-L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807. https://doi.org/10.1.1.44.7635
Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741. https://doi.org/10.1109/TPAMI.1984.4767596
Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov Chain Monte Carlo in practice*. Chapman & Hall.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.
Hinton, G. E., & Van Camp, D. (1993, July). *Keeping the neural networks simple by minimizing the description length of the weights* [Paper presentation]. Proceedings of the Sixth Annual Conference on Computational Learning Theory, Santa Cruz. https://doi.org/10.1145/168304.168306
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
Huang, P.-H., Chen, H., & Weng, L.-J. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika*, *82*, 329–354. https://doi.org/10.1007/s11336-017-9566-9
IEA. (2018). TIMSS 2019 student questionnaire (grade 8). International Association for the Evaluation of Educational Achievement. https://timssandpirls.bc.edu/timss2019/questionnaires/pdf/T19_StuQ_IntSc_8.pdf
Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: a Multidisciplinary Journal*, *23*, 555–566. https://doi.org/10.1080/10705511.2016.1154793
Khan, M. E. E., Bouchard, G., Murphy, K. P., & Marlin, B. M. (2010, December). *Variational bounds for mixed-data factor analysis* [Paper presentation]. Advances in Neural Information Processing Systems, 23, Vancouver, British Columbia, Canada.
Kozubowski, T. J., Podgórski, K., & Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*, *113*, 59–72. https://doi.org/10.1016/j.jmva.2012.02.010
Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, *5*, 369–411. https://doi.org/10.1214/10-BA607
Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, *51*, 519–539. https://doi.org/10.1080/00273171.2016.1168279
Meng, X.-L. (1994). Posterior predictive p-values. *Annals of Statistics*, *22*, 1142–1160.
Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. https://doi.org/10.1037/a0026802
Muthén, B., & Muthén, L. (2017). Mplus. In *Handbook of item response theory* (pp. 507–518). Chapman and Hall/CRC.
Neal, R. M., & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models* (pp. 355–368). Springer.
Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*, 681–686. https://doi.org/10.1198/016214508000000337
Pearl, J. (2009). *Causality*. Cambridge university press.
Saul, L., & Jordan, M. (1995, November). *Exploiting tractable substructures in intractable networks* [Paper presentation]. Advances in Neural Information Processing Systems, 8, Denver CO, USA.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Wang, Z., Gu, Y., Lan, A., & Baraniuk, R. (2020). Varfa: A variational factor analysis framework for efficient Bayesian learning analytics. arXiv preprint, arXiv:2005.13107. https://doi.org/10.48550/arXiv.2005.13107

Appendix A.

Full Conditional Distributions for Gibbs Sampler

For item j, denote $\tilde{\mathbf{a}}_j^{**}$ as the complement of $\mathbf{a}_j^{**}$ such that $\mathbf{a}_j^{**} \oplus \tilde{\mathbf{a}}_j^{**} = \mathbf{a}_j$ where $\oplus$ denotes a direct sum. Denote the corresponding part of $\mathbf{F}$ associated with $\mathbf{a}_j^{**}$ and $\tilde{\mathbf{a}}_j^{**}$ as $\mathbf{F}^{**}$ and $\tilde{\mathbf{F}}^{**}$, respectively. Denote $\boldsymbol{\nu}_{-j} = (\nu_{j1}, \cdots, \nu_{j,j-1}, \nu_{j,j+1}, \cdots, \nu_{jJ})^{\mathrm{T}}$ as the jth column vector of $\mathbf{V}$ without the diagonal term; $\mathbf{V}_{-jj}$ is denoted as the $(J-1) \times (J-1)$ submatrix of $\mathbf{V}$ without the jth row and column. Finally, set $\nu_{jj}^* = \nu_{jj} - \boldsymbol{\nu}_{-j}^T \mathbf{V}_{-jj}^{-1} \boldsymbol{\nu}_{-j}$. Note that $\nu_{jj}^* = \nu_{jj}$ when $\mathbf{V}$ is diagonal (i.e., no LD).

The posterior distribution for $\mathbf{a}_j^{**}$ conditional on everything else is:

$$\mathbf{a}_j^{**} | \text{rest} \sim N\left(\boldsymbol{\mu}_A^{**}, \boldsymbol{\Sigma}_A^{**}\right), \tag{A1}$$

with:

$$\boldsymbol{\Sigma}_A^{**} = \left[\boldsymbol{\Sigma}_{Aj}^{-1} + \nu_{jj}^{*-1} (\mathbf{F}^{**})' \mathbf{F}^{**}\right]^{-1}, \tag{A2}$$

and

$$\boldsymbol{\mu}_A^{**} = \boldsymbol{\Sigma}_A^{**} \left[(\mathbf{F}^{**})' \left(Y_j - \tilde{\mathbf{F}}^{**} \tilde{\mathbf{a}}_j^{**} - \mathbf{X}\mathbf{h}_j\right) + \boldsymbol{\mu}_{Aj} \boldsymbol{\Sigma}_{Aj}^{-1}\right]. \tag{A3}$$

Similarly, denote $\tilde{\mathbf{a}}_j^*$ as the complement of $\mathbf{a}_j^*$ such that $\mathbf{a}_j^* \oplus \tilde{\mathbf{a}}_j^* = \mathbf{a}_j$, with $\mathbf{F}^*$ and $\tilde{\mathbf{F}}^*$ as the corresponding part of $\mathbf{F}$ associated with $\mathbf{a}_j^*$ and $\tilde{\mathbf{a}}_j^*$, respectively. The posterior distribution for $\mathbf{a}_j^*$ is:

$$\mathbf{a}_j^* | \text{rest} \sim N\left(\boldsymbol{\mu}_A^*, \boldsymbol{\Sigma}_A^*\right), \tag{A4}$$

with:

$$\boldsymbol{\Sigma}_A^* = \left[\mathbf{T}_{Aj}^{-1} + \nu_{jj}^{*-1} (\mathbf{F}^*)' \mathbf{F}^*\right]^{-1}, \tag{A5}$$

$$\mathbf{T}_{Aj} = \text{diag}\left(\tau_{Aj1}^2, \ldots, \tau_{AjK_j^*}^2\right), \tag{A6}$$

and

$$\boldsymbol{\mu}_A^* = \boldsymbol{\Sigma}_A^* (\mathbf{F}^*)' \left(Y_j - \tilde{\mathbf{F}}^* \tilde{\mathbf{a}}_j^* - \mathbf{X}\mathbf{h}_j\right). \tag{A7}$$

The posterior distribution of τ_{Ajk}^{-2} follows the inverse Gaussian distribution:

$$\tau_{Ajk}^{-2}|\text{rest} \sim Inv - Gaussian\left(\mu' = \sqrt{\frac{\lambda_A^2 v_{jj}}{a_{jk}^2}}, \lambda' = \lambda_A^2\right), \tag{A8}$$

for $k = 1, \ldots, K_j^*$, and

$$\lambda_A^2|\text{rest} \sim Gamma\left(\alpha_A + \sum_{j=1}^{J} K_j^*, \beta_A + \sum_{j=1}^{J}\sum_{k=1}^{K_j^*} \tau_{Ajk}^2/2\right). \tag{A9}$$

The posterior distribution for $\mathbf{h}_j$ is:

$$\mathbf{h}_j|\text{rest} \sim N\left(\boldsymbol{\mu}_H^*, \boldsymbol{\Sigma}_H^*\right), \tag{A10}$$

with

$$\boldsymbol{\Sigma}_H^* = \left[\mathbf{T}_{Hj}^{-1} + v_{jj}^{*-1}\mathbf{X}\mathbf{X}'\right]^{-1}, \tag{A11}$$

$$\mathbf{T}_{Hj} = \text{diag}\left(\tau_{Hj1}^2, \ldots, \tau_{HjP}^2\right), \tag{A12}$$

and

$$\boldsymbol{\mu}_H^* = \boldsymbol{\Sigma}_H^*\mathbf{X}'(Y_j - \mathbf{F}\mathbf{a}_j). \tag{A13}$$

Meanwhile, the posterior distribution of τ_{Bjp}^{-2} follows the inverse Gaussian distribution:

$$\tau_{Hjp}^{-2}|\text{rest} \sim Inv - Gaussian\left(\mu' = \sqrt{\frac{\lambda_H^2 v_{jj}}{h_{jp}^2}}, \lambda' = \lambda_H^2\right), \tag{A14}$$

for$p = 1, \ldots, P$, and

$$\lambda_H^2|\text{rest} \sim Gamma\left(\alpha_H + JP, \beta_H + \sum_{j=1}^{J}\sum_{p=1}^{P} \tau_{Hjp}^2/2\right). \tag{A15}$$

When $\mathbf{V}$ is diagonal,

$$v_{jj}^{-1}|\text{rest} \sim Gamma\Big(\alpha_V + [N + K + P]/2, \beta_V + \left[\mathbf{S}_j + \left(\mathbf{a}_j^*\right)'\mathbf{T}_{Aj}\mathbf{a}_j^* + \mathbf{h}'_j\mathbf{T}_{Hj}\mathbf{h}_j\right]/2\Big), \tag{A16}$$

with $\mathbf{S}_j = \left(Y_j - \mathbf{F}\mathbf{a}_j - \mathbf{X}\mathbf{h}_j\right)'\left(Y_j - \mathbf{F}\mathbf{a}_j - \mathbf{X}\mathbf{h}_j\right)$. When $\mathbf{V}$ is a nondiagonal matrix with local dependence, an efficient block Gibbs sampler can be implemented by assigning double exponential priors to the off-diagonal elements, as shown in the PCFA literature (Chen, 2020, 2021b; Chen et al., 2021).

Moreover,

$$\mathbf{F} \sim N\left[\boldsymbol{\Sigma}_F^*(\mathbf{Y}\mathbf{V}^{-1}\mathbf{A} + \mathbf{X}\mathbf{B}'\mathbf{U}^{-1}), \boldsymbol{\Sigma}_F^*\right], \tag{A17}$$

with

$$\boldsymbol{\Sigma}_F^* = \left[\mathbf{U}^{-1} + \mathbf{A}'\mathbf{V}\mathbf{A}\right]^{-1}. \tag{A18}$$

On the structural level, the posterior distribution for $vec(\mathbf{B})$ can be entirely regularized:

$$vec(\mathbf{B})|\mathbf{U}, \text{rest} \sim N\left(\boldsymbol{\mu}_B^*, \mathbf{U}\otimes\boldsymbol{\Sigma}_B^*\right), \tag{A19}$$

with:

$$\boldsymbol{\Sigma}_B^* = \left[\mathbf{T}_B^{-1} + \mathbf{X}'\mathbf{X}\right]^{-1}, \tag{A20}$$

and

$$\boldsymbol{\mu}_B^* = \left(\boldsymbol{\Sigma}_B^*\mathbf{X}'\mathbf{F}\right)', \tag{A21}$$

with

$$\mathbf{T}_B = \text{diag}\left(\tau_{B1}^2, \ldots, \tau_{BP}^2\right). \tag{A22}$$

Meanwhile, the posterior distribution of τ_{Bp}^{-2} follows the inverse Gaussian distribution:

$$\tau_{Bp}^{-2}|\text{rest} \sim Inv - Gaussian\left(\mu' = \sqrt{\frac{\lambda_B^2}{\mathbf{b}_p'\mathbf{U}^{-1}\mathbf{b}_p}}, \lambda' = \lambda_B^2\right), \tag{A23}$$

for $p = 1, \ldots, P$, and

$$\lambda_B^2|\text{rest} \sim Gamma\left(\alpha_B + P, \beta_B + \sum_{p=1}^{P} \tau_{Bp}^2/2\right). \tag{A24}$$

$$\mathbf{U}^{-1}|\text{rest} \sim Wishart[\delta_U + N + P, \boldsymbol{\Sigma}_U + \left(\mathbf{F} - \mathbf{X}\mathbf{B}'\right)'\left(\mathbf{F} - \mathbf{X}\mathbf{B}'\right) + \mathbf{B}'\mathbf{T}_B\mathbf{B}], \tag{A25}$$

In Case 7, the disturbances are uncorrelated, which means there are only diagonal elements in $\mathbf{U}$:

$$u_{kk}^{-1}|\text{rest} \sim Gamma[\alpha_U + (N + P)/2, \beta_U + (F_k - \mathbf{X}\mathbf{b}_k{}')'(F_k - \mathbf{X}\mathbf{b}_k{}') + \mathbf{b}_k{}'\mathbf{T}_B\mathbf{b}_k], \tag{A26}$$

where $\mathbf{b}_k$ is the kth row of $\mathbf{B}$.

The posterior distribution for $\mathbf{B}^{**}$ is:

$$\mathbf{B}^{**}|\mathbf{U}, \text{rest} \sim N\left(\boldsymbol{\mu}_B^{**}, \mathbf{U}\otimes\boldsymbol{\Sigma}_B^{**}\right), \tag{A27}$$

with:

$$\boldsymbol{\Sigma}_B^{**} = \left[\boldsymbol{\Sigma}_B^{-1} + \mathbf{X}'\mathbf{X}\right]^{-1}, \tag{A28}$$

and

$$\boldsymbol{\mu}_B^{**} = \left[\boldsymbol{\Sigma}_B^{**}\left(\mathbf{X}'\mathbf{F} + \boldsymbol{\Sigma}_B^{-1}\boldsymbol{\mu}_B\right)\right]'. \tag{A29}$$

Meanwhile,

$$\mathbf{U}^{-1}|\text{rest} \sim Wishart[\delta_U + N + P, \boldsymbol{\Sigma}_U + \left(\mathbf{F} - \mathbf{X}\mathbf{B}'\right)'(\mathbf{F} - \mathbf{X}\mathbf{B}') + (\mathbf{B} - \boldsymbol{\mu}_B)\boldsymbol{\Sigma}_B(\mathbf{B} - \boldsymbol{\mu}_B)']. \tag{A30}$$

The highest posterior density (HPD) interval (Box & Tiao, 1973) is used to characterize the uncertainty of estimates. The convergence of the Markov chains is determined using the estimated potential scale reduction (EPSR) value (Gelman, 1996), which is the ratio of the weighted average of the between- and within-chain variance to the within-chain variance. When the ratio gets smaller than 1.1, the chains are usually considered convergent and stationary (Gelman et al., 2004). Model fit can be evaluated using the posterior predictive p or PP-p value (Gelman et al., 1996; Meng, 1994). Since the Bayesian Lasso was employed as a means for model selection, the PP-p value used here should be treated as complementary, which is usually considered acceptable when the value is not far away .5.

Appendix B.

Population Models in Simulation Studies

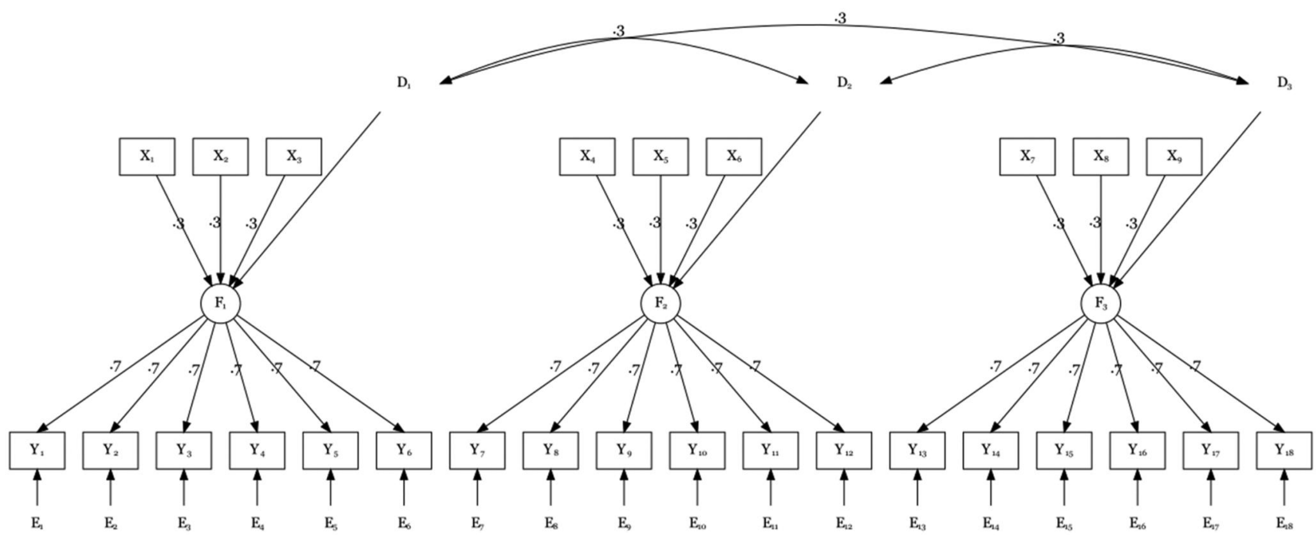


Figure B1. Population model without LD.

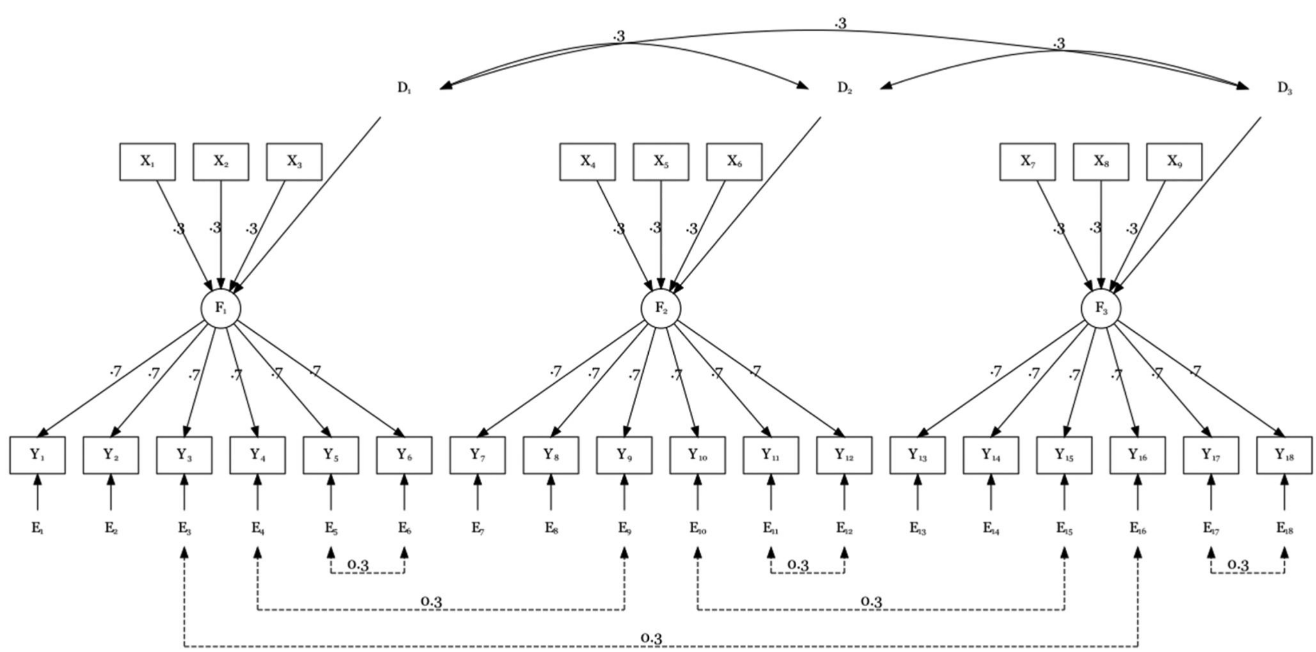


Figure B2. Population model with LD.

Appendix C

Table C1. Measurement part and structural part in TIMSS.

Item	Code	Content	Var	Code	Content
Y_1	BSBM16A	enjoy learning mathematics	X_1	BSBG03	often speak < lang of test > at home
Y_2	BSBM16B	wish have not to study math	X_2	BSBG04	amount of books in your home
Y_3	BSBM16C	math is boring	X_3	BSBG05A	home possess\computer tablet
Y_4	BSBM16D	learn interesting things	X_4	BSBG05B	home possess\study desk
Y_5	BSBM16E	like mathematics	X_5	BSBG05C	home possess\own room
Y_6	BSBM16F	like numbers	X_6	BSBG05D	home possess\internet connection
Y_7	BSBM16G	like math problems	X_7	BSBG05E	home possess\own mobile phone
Y_8	BSBM16H	look forward to math class	X_8	BSBG05F	home possess\<country specific>
Y_9	BSBM16I	favorite subject	X_9	BSBG05G	home possess\<country specific>
Y_{10}	BSBM19A	usually do well in math	X_{10}	BSBG05H	home possess\<country specific>
Y_{11}	BSBM19B	mathematics is more difficult	X_{11}	BSBG06A	highest lvl of edu of < parent a>
Y_{12}	BSBM19C	mathematics not my strength	X_{12}	BSBG06B	highest lvl of edu of < parent b>
Y_{13}	BSBM19D	learn quickly in mathematics	X_{13}	BSBG07	how far in edu do you expect to go
Y_{14}	BSBM19E	math makes nervous	X_{14}	BSBG08A	<parent a > born in < country>
Y_{15}	BSBM19F	good at working out problems	X_{15}	BSBG08B	<parent/ b > born in < country>
Y_{16}	BSBM19G	good at mathematics	X_{16}	BSBG10	how often absent from school
Y_{17}	BSBM19H	mathematics harder for me	X_{17}	BSBG11A	how often\tired
Y_{18}	BSBM19I	math makes confused	X_{18}	BSBG11B	how often\hungry
Y_{19}	BSBM20A	mathematics will help me	X_{19}	BSBM15	math\work on your own
Y_{20}	BSBM20B	need math to learn other things	X_{20}	BSBM26AA	math\how often teacher give homework
Y_{21}	BSBM20C	need math to get into university	X_{21}	BSBM26BA	math\how many minutes spent on homework
Y_{22}	BSBM20D	need math to get the job I want	X_{22}	BSBM27AA	math\extra lessons to excel last 12 month
Y_{23}	BSBM20E	job involving mathematics	X_{23}	BSBM27AA	math\extra lessons to keep up last 12 month
Y_{24}	BSBM20F	get ahead in the world	X_{24}	BSBM27BA	math\extra lessons how many month
Y_{25}	BSBM20G	more job opportunities	X_{25}	ITSEX	sex of students
Y_{26}	BSBM20H	parents think math important	X_{26}	BSDAGE	students age
Y_{27}	BSBM20I	important to do well in math			

Note. BSBM27AA is nominal and split into two variables.

Appendix D.

Significant Differential Coefficient Estimates in TIMSS

Table D1. Significant differential coefficients.

	X_2	X_3	X_4	X_5	X_{13}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{23}
Y_1									.044			
Y_2	−.047											
Y_3							−.035					
Y_4											−.040	
Y_7					−.032							
Y_8											−.030	.044
Y_9						.041	.030		−.032			
Y_{10}									**.050**		−.033	
Y_{12}									**−.059**			
Y_{13}										**.050**		.037
Y_{14}								**−.057**				
Y_{15}								.034	**.051**			−.042
Y_{18}				−.029			−.033					
Y_{19}					**.088**		−.039					**.058**
Y_{20}											−.042	.038
Y_{21}					−.076		.039	−.037				
Y_{22}	.037	−.033							**−.059**			
Y_{23}									**−.054**			**−.093**
Y_{24}					.045							
Y_{26}			.033		**−.074**	−.043						
Y_{27}												.041

Note. Absolute values above .05 are highlighted. Items and variables with no differential coefficients are not displayed.