Psychometrika (2025), **90**, 1236–1257
doi:10.1017/psy.2025.10027

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Bayesian Structural Equation Envelope Model

Chuchu Wang[1], Rongqian Sun[2], Xiangnan Feng[3] and Xinyuan Song[1]

[1]Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China; [2]School of Psychology, Shenzhen University, Shenzhen, China; [3]Department of Statistics and Data Science, Fudan University, Shanghai, China

Corresponding author: Xinyuan Song; Email: xysong@sta.cuhk.edu.hk

(Received 20 January 2025; revised 21 May 2025; accepted 22 May 2025; published online 8 August 2025)

A correction was made to the order of the authors from Rongqian Sun, Xiangnan Feng, Chuchu Wang and Xinyuan Song to Chuchu Wang, Rongqian Sun, Xiangnan Feng, Xinyuan Song. Chuchu Wang and Rongqian Sun are joint first authors.

Abstract

The envelope model has gained significant attention since its proposal, offering a fresh perspective on dimension reduction in multivariate regression models and improving estimation efficiency. One of its appealing features is its adaptability to diverse regression contexts. This article introduces the integration of envelope methods into the factor analysis model. In contrast to previous research primarily focused on the frequentist approach, the study proposes a Bayesian approach for estimation and envelope dimension selection. A Metropolis-within-Gibbs sampling algorithm is developed to draw posterior samples for Bayesian inference. A simulation study is conducted to illustrate the effectiveness of the proposed method. Additionally, the proposed methodology is applied to the ADNI dataset to explore the relationship between cognitive decline and the changes occurring in various brain regions. This empirical application further highlights the practical utility of the proposed model in real-world scenarios.

Keywords: Bayesian approach; envelope model; factor analysis; structural equation model

1. Introduction

The envelope model is initially introduced by Cook et al. (2010) as a technique for dimension reduction in multivariate analysis. Cook et al. (2010) first proposed the response envelope model, which aims to reduce the dimension of the response vector $\mathbf{Y}$. The core concept of the method is to decompose $\mathbf{Y}$ into a material and an immaterial part based on the assumption that certain linear combinations of $\mathbf{Y}$, known as X-invariant, remain unchanged regardless of variations in the predictor vector $\mathbf{X}$. These X-invariant linear combinations are considered immaterial to the regression and thus can reduce parameter space in estimation. Expanding upon this idea, Cook et al. (2013) extended the approach to include the dimensionality reduction in $\mathbf{X}$ within the context of multivariate regression, assuming the predictor $\mathbf{X}$ is stochastic. This extension, known as the predictor envelope model, is closely related to the partial least squares (PLS) method.

The envelope model has shown promising efficiency gains, which has prompted its extension into various contexts, including the partial envelope model (Su & Cook, 2011), generalized linear model (Cook & Zhang, 2015a), matrix-valued response model (Ding & Cook, 2018), sparse envelope model (Su et al., 2016) and spatial envelope model (Rekabdarkolaee et al., 2020). However, applying Bayesian methods to the envelope model has received limited attention compared to the frequentist framework in previous literature. This fact is primarily due to the challenge posed by parameterizing the envelope subspace within the Grassmann manifold space, where the basis of the envelope subspace is not unique.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Nevertheless, considering the envelope model from a Bayesian perspective is highly meaningful, as it allows for incorporating prior information into posterior inference without relying on asymptotic assumptions.

Regarding Bayesian envelope models, Khare et al. (2017) first proposed a Bayesian approach for analyzing response envelope models. They reparameterized the envelope model in a Steifel manifold to ensure the uniqueness of the orthogonal basis and adopted the Bingham distribution as a prior for the orthogonal basis matrix. However, this method is developed based on a specific design of the response envelope model, and it may lose parameter conjugacy in other scenarios. Additionally, the implementation of this method relies on sampling from the matrix Bingham distribution and the truncated inverse Gamma distribution, which can result in a heavy computational burden. On the other hand, Cook et al. (2016) developed a novel parameterization for the orthogonal basis matrix of the envelope subspace, and this idea was further extended to the Bayesian framework by Chakraborty & Su (2024). The proposed method does not rely on the Grassmann or Steifel manifold and can be applied to various envelope model contexts, including the response and predictor envelopes. Meanwhile, Lee et al. (2022) adopted the same technique and formulated the envelope model for Bayesian quantile regression.

Inspired by the successful application of the reparameterized envelope method, this study introduces a factor analytic technique to the predictor envelope model, enabling dimension reduction for both the response and predictor variables. Factor models, such as confirmatory factor analysis (CFA) and exploratory factor analysis (EFA), are widely utilized statistical tools in fields such as psychology, education, and social sciences. Over the past few decades, existing studies have demonstrated the effectiveness of the factor model in capturing latent structures and reducing dimensionality by summarizing the latent factors through multiple observed variables. Moreover, the joint modeling approach, also known as the structural equation model (SEM), has shown high potential for adapting to various modeling techniques. The joint model typically consists of two parts. The first part entails a factor analysis model aggregating multiple observed variables into latent factors. This step captures the underlying structure and interrelationships among the observed variables. The second part utilizes a regression model to elucidate the association between the latent factors and the observed covariates of interest. It allows for adapting various modeling techniques, enabling a flexible and versatile analysis. For example, Roy & Lin (2000) used multiple longitudinal measures as outcomes to quantify a latent variable of interest from different perspectives. They adopted a linear mixed model to study the effects of covariates on the time-dependent latent variable. Pan et al. (2019) integrated latent variables into a proportional hazards model to examine the observed and latent risk factors associated with the failure time of interest. Wang et al. (2021) introduced a latent-on-image model to jointly analyze high-dimensional imaging data and multiple clinical measurements in an Alzheimer's disease (AD) study. They characterized the severity of AD using various cognitive test scores as a latent factor in a CFA model and investigated the relationship between changes in brain structure and cognitive decline using a functional data regression model. Besides, researchers have developed plenty of methods to handle flexible data structure with a factor model, including hierarchical and heterogeneous data (Lee & Song, 2004), missing data (Song & Lee, 2006), longitudinal data (Song et al., 2011).

Bayesian methods have been extensively applied to SEM topics because they emphasize individual-level random observations and the estimation of first-order moment properties. This approach offers a simpler alternative to the traditional approach of fitting the covariance structure. By focusing on the raw individual-level data, Bayesian methods provide a flexible and intuitive framework for specifying prior distributions, incorporating prior knowledge, and conducting posterior inference. Furthermore, the hierarchical representation of the model, combined with efficient Markov chain Monte Carlo (MCMC) algorithms, allows for a straightforward statistical inference and accommodates highly complex models. For instance, Wang et al. (2016) integrated a mixture representation of the quantile regression model (Kozumi & Kobayashi, 2011; Yu & Moyeed, 2001) into SEM, moving beyond the usual assumption of normal errors. Feng et al. (2017) incorporated the Bayesian version of Lasso (BLasso) and adaptive Lasso (BaLasso) to quantile SEM, enabling simultaneous estimation and variable selection in this context.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

This study is motivated by a real-world data analysis of AD. A comprehensive understanding of the relationship between changes in brain structure, cognitive function, and the progression of brain disease is critical for accurate diagnosis and prevention of brain-related disorders. In the context of brain degeneration, it is often a global process that affects multiple regions rather than being confined to specific areas. Therefore, understanding the interdependencies among different regions of interest (ROIs) is crucial. The Envelope method, which we employed in this study, allows us to investigate and uncover these interdependencies among ROIs. By examining the relationships among various brain regions, we aim to identify the most informative ROI or combination of ROIs that can provide valuable insights into the underlying mechanisms of brain diseases.

The main contribution of this article is to formulate a novel envelope approach within the context of Bayesian SEM. Introducing the envelope model to the structural component of SEM serves a dual purpose. On the one hand, the envelope model eliminates immaterial variation in the data, thereby enhancing estimation efficiency. This enhancement is particularly valuable when confronted with high-dimensional candidate predictors (e.g., neuroimaging phenotypes derived from MRI data), of which only a small subset are genuinely influential so that high-dimensional predictors can be projected to a reduced subspace. On the other hand, grouping latent variables from multiple observed surrogates also represents a form of dimension reduction, especially pertinent in situations, where the correlation among multivariate responses arises from the shared underlying mechanism of reflecting the same latent construct from varying perspectives (e.g., cognitive impairment is manifested by multiple cognitive tests together in the ADNI study). Therefore, the proposed Bayesian Envelope SEM (BESEM) offers a novel perspective on concurrent dimension reduction for both response variables and predictors. This is achieved by employing factor analysis in the domain of latent variable modeling for the former and integrating an envelope structure into the structural equation for the latter. Even in scenarios, where a lower-dimensional envelope subspace does not exist, the proposed method seamlessly degenerates to a standard SEM without compromising parameter estimation accuracy. To our knowledge, this work is the first to introduce envelope methods to the SEM framework. We restrict the estimation of the envelope space to an orthogonal basis, which greatly reduces the computation efficiency. The derived posterior distributions are proved to be proper even with non-informative priors. A simple block Metropolis-within-Gibbs MCMC algorithm is presented to facilitate posterior sampling. The proposed Markov chain is shown to be ϕ-irreducible and aperiodic, ensuring the convergence of MCMC samples.

The remainder of the article is organized as follows. Section 2 outlines the envelope model and defines BESEM. Section 3 discusses the Bayesian inference. Section 4 presents simulated experiments, showing the efficiency gains of BESEM compared to conventional Bayesian SEM approaches. In Section 5, we apply BESEM to the Alzheimer's Disease Neuroimaging Initiative (ADNI) study to explore new insights into the relationship between cognitive decline and different brain regions. Technical details are provided in the Supplementary Material.

2. Model description

2.1. Review of envelope model

In this section, we provide a brief overview of the envelope idea proposed by Cook et al. (2010). This idea was initially developed to reduce the regression coefficients in the multivariate linear regression model given by

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^r$, $\mathbf{X} \in \mathbb{R}^p$ represent the response and predictor vector, respectively, $\boldsymbol{\mu} \in \mathbb{R}^r$ and $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$ are the intercept and coefficients, and $\boldsymbol{\varepsilon} \in \mathbb{R}^r$ is the error term with a zero mean and a positive definite covariance matrix $\boldsymbol{\Sigma}$.

The response envelope model aims to decompose the response variable $\mathbf{Y}$ into a material part and an immaterial part based on the assumption that X-invariant linear combinations of $\mathbf{Y}$ exist. Specifically,

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

let $\mathcal{E}$ be a subspace of $\mathbb{R}^r$, $\mathbf{P}_{\mathcal{E}}$ is the projection onto $\mathcal{E}$ and $\mathbf{Q}_{\mathcal{E}} = \mathbf{I}_r - \mathbf{P}_{\mathcal{E}}$, where $\mathbf{I}_r$ is the identity matrix of size r. The response envelope seeks to find the minimal subspace $\mathcal{E}$ that satisfies the following two conditions:

(i) $\mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X} \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y}$, where $\sim$ denotes equality in distribution and
(ii) $\mathbf{P}_{\mathcal{E}}\mathbf{Y} \perp\!\!\!\perp \mathbf{Q}_{\mathcal{E}}\mathbf{Y} \mid \mathbf{X}$.

These two conditions stipulate that the distribution of $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ is not influenced by $\mathbf{X}$ nor by $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$, indicating that the dependence of $\mathbf{Y}$ on $\mathbf{X}$ is concentrated only in $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$, which is referred to as the material part. Then, $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ is the immaterial part. Cook et al. (2010) showed that (i) and (ii) are equivalent to the following two conditions:

(i′) $\mathcal{B} \subseteq \mathcal{E}$, where $\mathcal{B} = \text{span}(\boldsymbol{\beta})$ and
(ii′) $\boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{E}}$.

Condition (ii′) states that $\mathcal{E}$ is a reducing subspace of $\boldsymbol{\Sigma}$. Combined with condition (i′), the $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, is defined as the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\text{span}(\boldsymbol{\beta})$. The existence of the $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ was discussed by Cook et al. (2010).

Let $u = \dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}))$, $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times u}$ and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r\times(p-u)}$ be the orthogonal bases of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ and $\mathcal{E}^{\perp}_{\boldsymbol{\Sigma}}(\mathcal{B})$, respectively. Here, $\mathcal{E}^{\perp}_{\boldsymbol{\Sigma}}(\mathcal{B})$ is the orthogonal complement of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$. Model (1) can be parameterized in terms of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$ as follows:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{d}\mathbf{X} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{2}$$

where $\boldsymbol{d} \in \mathbb{R}^{u\times p}$ is the coordinates of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$, and we have $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{d}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u\times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u)\times(r-u)}$ are positive definite and specify the covariance structure of the material and immaterial parts. For a fixed u, the total number of parameters required for the model (2) is

$$\begin{aligned} & r + pu + u(r-u) + \frac{u(u+1)}{2} + \frac{(r-u)(r-u+1)}{2} \\ = \; & r + pu + \frac{r(r+1)}{2}. \end{aligned}$$

If $u < r$, the efficiency gains of the envelope are possible compared to the standard multivariate linear regression model (1). If $u = r$, the envelope model degenerates into the linear regression standard model.

The predictor envelope model is built on a framework similar to the response envelope model, aiming to reduce the dimensionality of $\mathbf{X}$. To accommodate a stochastic predictor $\mathbf{X}$ with mean $\boldsymbol{\mu}_{\mathbf{X}}$ and variance $\boldsymbol{\Sigma}_{\mathbf{X}}$, model (1) is modified as follows:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\beta}^T\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right) + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$ and not necessarily normally distributed. To decompose $\mathbf{X}$ into its material and immaterial parts, we assume there is a subspace $\mathcal{S} \in \mathbb{R}^p$ that satisfies the following conditions:

(a) $\text{cov}(\mathbf{Y}, \mathbf{Q}_{\mathcal{S}}\mathbf{X} \mid \mathbf{P}_{\mathcal{S}}\mathbf{X}) = 0$ and
(b) $\text{cov}(\mathbf{P}_{\mathcal{S}}\mathbf{X}, \mathbf{Q}_{\mathcal{S}}\mathbf{X}) = 0$.

These conditions indicate that $\mathbf{Q}_{\mathcal{S}}\mathbf{X}$ is uncorrelated with both $\mathbf{P}_{\mathcal{S}}\mathbf{X}$ and $\mathbf{Y}$, while all the information in $\mathbf{X}$ that is linearly related to the regression is captured by $\mathbf{P}_{\mathcal{S}}\mathbf{X}$. It has been shown that these conditions are equivalent to the following ones (Cook et al., 2013):

(a′) $\mathcal{B}' \in \mathcal{S}$, and
(b′) $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}_{\mathbf{X}}$,

where $\mathcal{B}' = \text{span}(\boldsymbol{\beta}^T)$. The intersection of all subspaces satisfying these two properties is referred to as the $\boldsymbol{\Sigma}_{\mathbf{X}}$-envelope of $\mathcal{B}'$, denoted as $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\mathcal{B}')$. Let $m = \dim\left(\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\mathcal{B}')\right)$, and $\boldsymbol{\Phi} \in \mathbb{R}^{p\times m}$ be a orthogonal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_{\mathbf{X}}}(\mathcal{B}')$. Then, the predictor envelope model is formulated as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{c}^T\boldsymbol{\Phi}^T\left(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}\right) + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\Sigma}_{\mathbf{X}} = \boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\Phi}^T + \boldsymbol{\Phi}_0\boldsymbol{\Delta}_0\boldsymbol{\Phi}_0^T, \tag{4}$$

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

where $\boldsymbol{\beta} = \boldsymbol{\Phi}\boldsymbol{c}$, $\boldsymbol{c} \in \mathbb{R}^{m\times r}$ is the coordinate of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Phi}$, $\Phi_0 \in \mathbb{R}^{p\times(p-m)}$ is an orthogonal basis of $\mathcal{E}^{\perp}_{\boldsymbol{\Sigma}_X}(\mathcal{B}')$, and $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}_0$ are positive definite matrices.

Considering that reducing the dimensionality of complex covariates is often of great interest in efficiently quantifying the relationship between the latent outcomes and a set of predictors in SEM, in the following section, we focus on the aspect of predictor dimension reduction to describe the proposed methodology.

2.2. BESEM

We introduce the envelope approach in the context of SEM. Let $\mathbf{Y} \in \mathbb{R}^r$ be the observed response vector like the previous representation. In addition, we let $\mathbf{X} \in \mathbb{R}^p$ be the vector of random predictor variables with mean $\boldsymbol{\mu}_X$ and variance $\boldsymbol{\Sigma}_X$, and $\boldsymbol{\eta} \in \mathbb{R}^q$ be a vector of latent variables that are expected to be formulated from the observed variables in $\mathbf{Y}$. A standard SEM can be defined as follows:

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu}_Y + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\eta} &= \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\beta}^T(\mathbf{X} - \boldsymbol{\mu}_X) + \boldsymbol{\delta}, \end{aligned} \tag{5}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{r\times q}$ is the unknown factor loading matrix, $\boldsymbol{\mu}_Y \in \mathbb{R}^r$ and $\boldsymbol{\mu}_{\boldsymbol{\eta}} \in \mathbb{R}^q$ are intercepts, $\boldsymbol{\beta} \in \mathbb{R}^{p\times q}$ is the unknown coefficient of interest, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are independent error terms. We assume $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ and $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\delta)$, where $\boldsymbol{\Sigma}_\varepsilon$ is a diagonal matrix with diagonal elements $\left\{\sigma^2_{\varepsilon k}\right\}^r_{k=1}$, and $\boldsymbol{\Sigma}_\delta$ is a positive definite matrix.

The first equation in (5) represents the link between the observed outcome variables $\mathbf{Y}$ and the latent variables $\boldsymbol{\eta}$, characterized by a CFA model. The second equation in (5) is a structural equation to assess the effects of the covariates of interest $\mathbf{X}$ on $\boldsymbol{\eta}$.

To reduce the dimensionality of $\mathbf{X}$ with an envelope structure and decompose $\mathbf{X}$ into its material and immaterial parts, we assume there is a subspace $\mathcal{S} \in \mathbb{R}^p$ that satisfies the following conditions:

(a) $\mathrm{cov}(\boldsymbol{\eta}, \mathbf{Q}_{\mathcal{S}}\mathbf{X} \mid \mathbf{P}_{\mathcal{S}}\mathbf{X}) = 0$ and
(b) $\mathrm{cov}(\mathbf{P}_{\mathcal{S}}\mathbf{X}, \mathbf{Q}_{\mathcal{S}}\mathbf{X}) = 0$.

These conditions indicate that $\mathbf{Q}_{\mathcal{S}}\mathbf{X}$ is uncorrelated with both $\mathbf{P}_{\mathcal{S}}\mathbf{X}$ and $\boldsymbol{\eta}$, while all the information in $\mathbf{X}$ that is linearly related to $\boldsymbol{\eta}$ in the SEM is captured by $\mathbf{P}_{\mathcal{S}}\mathbf{X}$. The conditions (a) and (b) are equivalent to the following two conditions (Cook et al., 2013):

(a′) $\mathcal{B}' \in \mathcal{S}$, where $\mathcal{B}' = \mathrm{span}(\boldsymbol{\beta}^T)$ and
(b′) $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}_X$.

Therefore, we aim to find the smallest subspace that satisfies (a′) and (b′), or equivalently, the intersection of all reducing subspaces of $\boldsymbol{\Sigma}_X$ that contains $\mathcal{B}'$, which is the $\boldsymbol{\Sigma}_X$-envelope of $\mathcal{B}'$, denoted as $\mathcal{E}_{\boldsymbol{\Sigma}_X}(\mathcal{B}')$.

Let $m = \dim\left(\mathcal{E}_{\boldsymbol{\Sigma}_X}(\mathcal{B}')\right)$, and $\boldsymbol{\Phi} \in \mathbb{R}^{p\times m}$ be a orthogonal basis of $\mathcal{E}_{\boldsymbol{\Sigma}_X}(\mathcal{B}')$. Then, the SEM model (5) can be formulated as the following envelope SEM:

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu}_Y + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\eta} &= \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{c}^T\boldsymbol{\Phi}^T\left(\mathbf{X} - \boldsymbol{\mu}_X\right) + \boldsymbol{\delta}, \\ \boldsymbol{\Sigma}_X &= \boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\Phi}^T + \boldsymbol{\Phi}_0\boldsymbol{\Delta}_0\boldsymbol{\Phi}_0^T, \end{aligned} \tag{6}$$

where $\boldsymbol{\beta}^T = \boldsymbol{\Phi}\boldsymbol{c}$, $\boldsymbol{c} \in \mathbb{R}^{m\times q}$ is the full rank coordinate of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Phi}$, $\boldsymbol{\Phi}_0 \in \mathbb{R}^{p\times(p-m)}$ is an orthogonal basis of $\mathcal{E}^{\perp}_{\boldsymbol{\Sigma}_X}(\mathcal{B}')$. The matrices $\boldsymbol{\Delta} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{\Delta}_0 \in \mathbb{R}^{(p-m)\times(p-m)}$ are positive definite. When $m = p$, the proposed envelope model is equivalent to a standard SEM.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

2.3. Reparametrization of BESEM

For a fixed m, the envelope $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$ is defined on a $p \times m$ Grassmann manifold, which implies that the specification of the basis $\mathbf{\Phi}$ is not unique. We can reparameterize the envelope model to address this issue and identify a unique orthogonal basis in an Euclidean space (Cook et al., 2016).

Let $\mathbf{\Phi}$ be an arbitrary basis of $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$. We define $\mathbf{\Phi}_1$ as the first m rows of $\mathbf{\Phi}$. Without loss of generality, we assume that $\mathbf{\Phi}_1$ is nonsingular (otherwise, we can reorder the rows in $\mathbf{X}$). The remaining $p-m$ rows of $\mathbf{\Phi}$ are denoted as $\mathbf{\Phi}_2$. $\mathbf{\Phi}$ can be expressed as follows:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{\Phi}_2\mathbf{\Phi}_1^{-1} \end{bmatrix} \mathbf{\Phi}_1 = \begin{bmatrix} \mathbf{I}_m \\ \boldsymbol{A} \end{bmatrix} \mathbf{\Phi}_1 = \boldsymbol{C}_A \mathbf{\Phi}_1, \tag{7}$$

where $\boldsymbol{A} = \mathbf{\Phi}_2\mathbf{\Phi}_1^{-1} \in \mathbb{R}^{(p-m)\times m}$ is an unconstrained matrix, and $\boldsymbol{C}_A = \left(\mathbf{I}_m, \boldsymbol{A}^T\right)^T$ is also a basis of $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$. It has been shown that $\boldsymbol{A}$ depends on $\mathbf{\Phi}$ only through span($\mathbf{\Phi}$): suppose $\mathbf{\Phi}^*$ is a different basis of $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$ and $\boldsymbol{O} \in \mathbb{R}^{m\times m}$ is a full rank matrix, such that $\mathbf{\Phi}^* = \mathbf{\Phi}\boldsymbol{O}$, then $\mathbf{\Phi}_1^* = \mathbf{\Phi}_1\boldsymbol{O}$, $\mathbf{\Phi}_2^* = \mathbf{\Phi}_2\boldsymbol{O}$, and $\boldsymbol{A}^* = \mathbf{\Phi}_2\boldsymbol{O}\boldsymbol{O}^{-1}\mathbf{\Phi}_1 = \boldsymbol{A}$ (Su et al., 2016). Therefore, there is a one-to-one correspondence between $\boldsymbol{A} \in \mathbb{R}^{(p-m)\times m}$ and $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$. Furthermore, the orthonormal basis of $\mathcal{E}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$ can be expressed as $\mathbf{\Phi} = \mathbf{\Phi}(\boldsymbol{A}) = \boldsymbol{C}_A\left(\boldsymbol{C}_A^T\boldsymbol{C}_A\right)^{-\frac{1}{2}}$.

Based on (Chen et al., 2020, Lemma 1), we can construct a basis of $\mathcal{E}^{\perp}_{\mathbf{\Sigma}_\mathbf{X}}(\mathcal{B}')$ by forming the matrix $\boldsymbol{D}_A = (-\boldsymbol{A}, \mathbf{I}_{p-m})^T$. Likewise, we can obtain an orthonormal basis as $\mathbf{\Phi}_0(\boldsymbol{A}) = \boldsymbol{D}_A(\boldsymbol{D}_A^T\boldsymbol{D}_A)^{-\frac{1}{2}}$. Accordingly, (5) can be written as

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu}_\mathbf{Y} + \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\eta} &= \boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{c}^T\mathbf{\Phi}(\boldsymbol{A})^T\left(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}\right) + \boldsymbol{\delta}, \end{aligned} \tag{8}$$

where $\mathbf{X} \sim N\left(\boldsymbol{\mu}_\mathbf{X}, \mathbf{\Phi}(\boldsymbol{A})\mathbf{\Delta}\mathbf{\Phi}(\boldsymbol{A})^T + \mathbf{\Phi}_0(\boldsymbol{A})\mathbf{\Delta}_0\mathbf{\Phi}_0(\boldsymbol{A})^T\right)$.

2.4. Model identification

The measurement equation in (8) is not identified because, for any nonsigular matrix $\mathbf{B}$, we have $\mathbf{\Lambda}\boldsymbol{\eta} = \mathbf{\Lambda}\mathbf{B}\mathbf{B}^{-1}\boldsymbol{\eta} = \mathbf{\Lambda}^*\boldsymbol{\eta}^*$, where $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{B}$, and $\boldsymbol{\eta}^* = \mathbf{B}^{-1}\boldsymbol{\eta}$ is still random latent variables. Imposing identification constraints on the measurement equation is necessary to establish identification. One commonly used approach, as described in Song & Lee (2012), is to define $\mathbf{\Lambda}$ in a non-overlapping structure with a fixed nonzero element in each column. Here is an illustrative example: consider a scenario with $p = 6$ observed variables and $q = 2$ latent variables. In this case, the first four observed variables are associated with the first latent factor, while the remaining two are related to the second latent factor. Let λ_{jk} denote the (j,k)th element of $\mathbf{\Lambda}$. A non-overlapping structure of $\mathbf{\Lambda}$ can be defined as follows:

$$\mathbf{\Lambda}^T = \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & \lambda_{41} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{62} \end{bmatrix}.$$

In the above structure, the elements with values 1 and 0 are known parameters with fixed values, while the λ_{jk}s' represent the unknown parameters that need to be estimated. Therefore, the total number of parameters that need to be estimated is

$$\begin{aligned} K(m) &= r + (r-q) + r + q + mq + p + (p-m)m \\ &\quad + \frac{m(m+1)}{2} + \frac{(p-m)(p-m+1)}{2} + \frac{q(q+1)}{2} \\ &= 3r + \frac{q(q+2m+1)}{2} + \frac{p(p+3)}{2}. \end{aligned}$$

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

2.5. Envelope dimension selection

Determining the optimal dimension of the envelope space is an essential step before estimating the envelope model. Various information criteria, such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), and the deviance information criteria(DIC), have been widely used in previous research. Information criteria offer a trade-off between model fit and complexity, with lower values indicating better-fitting models. The dimension of the envelope space that minimizes the AIC, BIC, or DIC can be chosen as the optimal dimension. However, there is no universally applicable guideline regarding which information criterion performs best in all scenarios. The performance of these criteria may vary depending on the specific characteristics of the data and the model under consideration. Prior studies, such as Shen et al. (2023) and Khare et al. (2017); and Chakraborty & Su (2024), have demonstrated that different information criteria have distinct performance in various envelope model contexts. Therefore, we investigate the performance of several information criteria, including AIC, BIC, DIC, and average weighted estimation (AWE), in determining the optimal dimension of the envelope space in the context of BESEM by conducting an empirical experiment in Section 4. Let $\boldsymbol{\theta}$ be all the unknown parameters and $L(\text{data}|\boldsymbol{\theta})$ be the likelihood, then $\hat{L}$ is the maximized likelihood value of the model. The deviance is $D(\boldsymbol{\theta}) = -2\log L(\text{data}|\boldsymbol{\theta})$. The values of AIC, BIC, DIC, and AWE are calculated as follows:

$$\begin{aligned}
\text{AIC}_m &= -2\cdot\log\hat{L} + 2\cdot K(m),\\
\text{BIC}_m &= -2\cdot\log\hat{L} + \log n\cdot K(m),\\
\text{DIC}_m &= D(\bar{\boldsymbol{\theta}}) + 2p_D,\\
\text{AWE}_m &= -2\cdot\log\hat{L} + 2\cdot K(m)\cdot\left(\frac{3}{2} + \log n\right),
\end{aligned}$$

where $p_D = E_{\boldsymbol{\theta}|\text{data}}[D] - D(E_{\boldsymbol{\theta}|\text{data}}[\boldsymbol{\theta}]) = \bar{D} - D(\bar{\boldsymbol{\theta}})$ denotes the posterior mean deviance minus the deviance evaluated at the posterior mean of the parameters.

This empirical analysis provides valuable insights into the relative performance of information criteria and their suitability in specific scenarios of envelope models.

3. Bayesian inference

3.1. Prior distributions

Let $\boldsymbol{\theta}$ represent all the unknown parameters in model (8), and $\boldsymbol{\theta} = \{\boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_\delta, \boldsymbol{\mu}_{\boldsymbol{\eta}}, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Delta}, \boldsymbol{\Delta}_0, \boldsymbol{c}, \boldsymbol{A}\}$. We define a joint prior density $p(\boldsymbol{\theta})$ for the unknown parameters, which can be decomposed as $p(\boldsymbol{\theta}) = p(\boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\Lambda} \mid \boldsymbol{\Sigma}_\varepsilon)p(\boldsymbol{\Sigma}_\varepsilon)p(\boldsymbol{\Sigma}_\delta)p(\boldsymbol{\mu}_{\boldsymbol{\eta}})p(\boldsymbol{\mu}_{\mathbf{X}})p(\boldsymbol{\Delta})p(\boldsymbol{\Delta}_0)p(\boldsymbol{c}|\boldsymbol{A}, \boldsymbol{\Sigma}_\delta)p(\boldsymbol{A})$.

We first specify some notations: $\mathbb{S}_+^{m\times m}$ denotes the set of $m\times m$ symmetric positive definite matrices, $\mathbb{R}_+^m$ denotes the set of vectors of length m with positive values, $\text{IW}_m(\boldsymbol{\Psi}, \nu)$ represents the Inverse-Wishart distribution with scale matrix $\boldsymbol{\Psi} \in \mathbb{S}_+^{m\times m}$ and degree of freedom ν, $\text{IG}(a, b)$ represents the Inverse-Gamma distribution, and $\mathcal{MN}_{n_1,n_2}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ is the matrix normal distribution with the parameters $\mathbf{M} \in \mathbb{R}^{n_1\times n_2}$, $\mathbf{U} \in \mathbb{S}_+^{n_1\times n_1}$, $\mathbf{V} \in \mathbb{S}_+^{n_2\times n_2}$.

The priors for the concerned parameters are defined as follows:

- Let $\boldsymbol{\Lambda}_y = \left(\boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\Lambda}\right)$, then for each $k = 1, \ldots, r$, where $\boldsymbol{\Lambda}_{yk}^T$ represents the kth row of $\boldsymbol{\Lambda}_y$, we assign a joint prior to $(\boldsymbol{\Lambda}_{yk}, \sigma_{\varepsilon k}^2)$, given by $p(\boldsymbol{\Lambda}_{yk}, \sigma_{\varepsilon k}^2) = p(\boldsymbol{\Lambda}_{yk}|\sigma_{\varepsilon k}^2)p(\sigma_{\varepsilon k}^2)$. Specifically,

$$\left[\boldsymbol{\Lambda}_{yk} \mid \sigma_{\varepsilon k}^2\right] \sim \text{N}_{q+1}(\boldsymbol{\Lambda}_{0k}, \mathbf{H}_{0k}), \qquad \sigma_{\varepsilon k}^2 \sim \text{IG}(a_{0k}, b_{0k}), \quad \text{for } k = 1, \ldots, r,$$

  where $\boldsymbol{\Lambda}_{0k} \in \mathbb{R}^{q+1}$, $\mathbf{H}_{0k} \in \mathbb{S}_+^{(q+1)\times(q+1)}$, $a_{0k} > 0$ and $b_{0k} > 0$ are prefixed hyperparameters.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

- We adopt flat priors for the intercept terms $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\mu}_{\boldsymbol{\eta}}$

$$p(\boldsymbol{\mu}_{\mathbf{X}}) \propto 1, \qquad p(\boldsymbol{\mu}_{\boldsymbol{\eta}}) \propto 1.$$

- We assign the Inverse-Wishart distribution to $\boldsymbol{\Delta}$, $\boldsymbol{\Delta}_0$, and $\boldsymbol{\Sigma}_\delta$:

$$\boldsymbol{\Delta} \sim \mathrm{IW}_m(\boldsymbol{\Psi}_1, \nu_1), \qquad \boldsymbol{\Delta}_0 \sim \mathrm{IW}_{p-m}(\boldsymbol{\Psi}_2, \nu_2), \qquad \boldsymbol{\Sigma}_\delta \sim \mathrm{IW}_q(\boldsymbol{\Psi}_\delta, \nu_0),$$

  where $\boldsymbol{\Psi}_1 \in \mathbb{S}_+^{m\times m}$, $\boldsymbol{\Psi}_2 \in \mathbb{S}_+^{(p-m)\times(p-m)}$, $\boldsymbol{\Psi}_\delta \in \mathbb{S}_+^{q\times q}$, $\nu_1 > m-1$, $\nu_2 > p-m-1$, and $\nu_0 > q-1$, are hyperparameters.
- We use matrix normal prior for $\boldsymbol{c}$ and $\boldsymbol{A}$:

$$[\boldsymbol{c} \mid \boldsymbol{A}, \boldsymbol{\Sigma}_\delta] \sim \mathcal{MN}_{m,q}\left(\mathbf{M}^{-1}\boldsymbol{\Phi}(\boldsymbol{A})^T\boldsymbol{e}, \mathbf{M}, \boldsymbol{\Sigma}_\delta\right),$$
$$\boldsymbol{A} \sim \mathcal{MN}_{p-m,m}(\boldsymbol{A}_0, \boldsymbol{K}, \boldsymbol{L}),$$

  where $\mathbf{M} \in \mathbb{S}_+^{m\times m}$, $\boldsymbol{e} \in \mathbb{R}^{p\times q}$, $\boldsymbol{A}_0 \in \mathbb{R}^{(p-m)\times m}$, $\boldsymbol{K} \in \mathbb{S}_+^{(p-m)\times(p-m)}$, $\boldsymbol{L} \in \mathbb{S}_+^{m\times m}$ are hyperparameters.

Vague (noninformative) priors are used in the numerical studies. Such vague priors are appropriate choices widely adopted when limited information is available about the relationship between the latent responses and a large number of covariates. However, the proposed method can also straightforwardly accommodate prior knowledge once available. For instance, if we have prior information on the factor loading matrix $\boldsymbol{\Lambda}$ in the form of $\widehat{\boldsymbol{\Lambda}}_{prior}$, we can set the prior mean of the loading matrix as $\boldsymbol{\Lambda}_0 = \widehat{\boldsymbol{\Lambda}}_{prior}$ with a relatively small prior variance matrix $\mathbf{H}_{0k}$. Similarly, if we possess prior knowledge on the potential envelope subspace, i.e., $\widehat{\mathcal{E}}_{prior}$, we can determine the prior information for $\mathbf{A}$, $\widehat{\mathbf{A}}_{prior}$, through the one-to-one correspondence in Equation (7) and set $\mathbf{A}_0 = \widehat{\mathbf{A}}_{prior}$. In line with the methodology proposed by Chakraborty & Su (2024), even partial prior information concerning $\widehat{\mathcal{E}}_{prior}$ can be incorporated into the proposed method. For example, assuming a candidate envelope dimension of $m = 4$ and we only know about $\widehat{\mathcal{E}}_{prior}$ that it contains two independent unit vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. In such circumstances, we can generate two random vectors from $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2)^\perp$ as $(\mathbf{v}_3, \mathbf{v}_4) = \mathbf{G}_0\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1/2}$, where $\mathbf{G}_0 \in \mathbb{R}^{p\times(p-2)}$ is an orthonormal basis of $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2)^\perp$ and $\mathbf{C}$ is a $(p-2)\times p$ matrix with each entry independently generated from the standard normal distribution. Subsequently, the prior orthonormal basis can be formulated as $\widehat{\boldsymbol{\Phi}}(\mathbf{A})_{prior} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$ and $\widehat{\mathbf{A}}_{prior}$ can be derived through Equation (7).

3.2. Posterior analysis and sampling process

Let $\mathcal{D} = \{\mathbb{X}, \mathbb{Y}\}$ represent the collection of n independent observations of $(\mathbf{X}, \mathbf{Y})$, where $\mathbb{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\mathbb{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$. Additionally, let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$ denote the matrix of latent variables. The Bayesian estimate of $\boldsymbol{\theta}$ is commonly defined as the sample mean or mode of the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$. However, drawing samples from $p(\boldsymbol{\theta}|\mathcal{D})$ can be challenging due to the presence of the latent variable $\boldsymbol{\eta}$, as $p(\boldsymbol{\theta}|\mathcal{D})$ may not have a closed form. We utilize the data augmentation technique proposed by Tanner & Wong (1987) to address this issue. In this approach, we treat the latent variables $\boldsymbol{\eta}$ as missing data and augment the observed data with them. Consequently, the posterior sampling procedure can be constructed based on the complete data set and the joint distribution of $[\boldsymbol{\theta}, \boldsymbol{\eta}|\mathcal{D}]$. Theorem 1 establishes the propriety of the posterior density, and its proof is included in the Supplementary Material.

Theorem 1 (Posterior propriety). *The posterior density $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathcal{D})$ is proper with respect to Lebesgue measure on $\mathbb{R}^r \times \mathbb{R}^{r-q} \times \mathbb{R}_+^r \times \mathbb{S}_+^{q\times q} \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{S}_+^{m\times m} \times \mathbb{S}_+^{(p-m)\times(p-m)} \times \mathbb{R}^{m\times q} \times \mathbb{R}^{(p-m)\times m} \times \mathbb{R}^q$.*

Given the complexity of the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathcal{D})$ and the challenge of directly sampling from it, we propose a Metropolis-within-Gibbs sampler to draw posterior samples. The Gibbs sampler allows us to generate samples for each element of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ from their respective full conditional distributions iteratively. The proposed MCMC sampler (Algorithm 1) for the case when envelope dimension $m \in \{1, \ldots, p-1\}$ can be found in the Supplementary Material.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

We also prove that the Markov chain generated by the proposed algorithm is ϕ-irreducible and aperiodic in Theorem 2, which ensures the convergence to its stationary distribution from almost all starting states (Cinlar, 2013).

Theorem 2 (ϕ-irreducible and aperiodic). *The Markov chain generated by the Metropolis-within-Gibbs algorithm for the posterior sampling is ϕ-irreducible and aperiodic.*

4. Simulation

We design a simulation experiment to assess the performance of the BESEM methodology proposed in this article. The data are generated using equations (7) and (8). We consider three dimensions of the underlying envelope space, denoted as m_{true}, which include values of 2, 4, and 6. Additionally, we consider two different settings for the predictor dimension p, with values of 20 and 40. The other parameters are fixed as follows: $r = 6$ and $q = 2$. To ensure identification, $\lambda_{1,1}$ and $\lambda_{4,2}$ are set to 1, while the remaining free elements of $\mathbf{\Lambda}$ are generated independently from a uniform distribution Unif$(-5,5)$. The elements of $\boldsymbol{\mu}_{\boldsymbol{\eta}}$, $\boldsymbol{\mu}_{\mathbf{X}}$, and $\boldsymbol{\mu}_{\mathbf{Y}}$ are independently sampled from a uniform distribution Unif$(-10,10)$, and the elements of $\boldsymbol{c}$ and $\boldsymbol{A}$ are drawn independently from uniform distributions Unif$(5,10)$ and Unif$(0,5)$, respectively. The matrices $\mathbf{\Delta}, \mathbf{\Delta}_0$ and $\mathbf{\Sigma}_\delta$ are generated independently from IW$_m(500\mathbf{I}_m, m+2)$, IW$_{p-m}(\mathbf{I}_{p-m}, p-m+2)$, and IW$_q(10\mathbf{I}_q, q+2)$, respectively. The diagonal elements of the error covariance matrix $\mathbf{\Sigma}_\varepsilon$ are simulated from Unif$(0.5,5)$. More specifically, we first generate $\mathbf{X}$ from a multivariate normal distribution characterized by parameters $\boldsymbol{\mu}_{\mathbf{X}}$, $\boldsymbol{A}$, $\mathbf{\Delta}$, and $\mathbf{\Delta}_0$ as mentioned before in Section 2.3. The error terms $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are independently drawn from normal distributions with zero mean and covariance matrix $\mathbf{\Sigma}_\delta$ and $\mathbf{\Sigma}_\varepsilon$, respectively. The latent variables $\boldsymbol{\eta}$ and the corresponding observed responses $\mathbf{Y}$ are then generated according to Equation (8). The simulation experiment considers different sample sizes, namely, 50, 150, 300, and 600. For each sample size, 100 replicated datasets are generated.

For model estimation and Bayesian inference, we consider vague priors on the unknown parameters as follows:

(1) For the prior distributions of $\mathbf{\Lambda}_{yk}$ and $\sigma^2_{\varepsilon k}$, we set $\mathbf{\Lambda}_{0k} = \mathbf{0}$, $\mathbf{H}_{0k} = 1000\mathbf{I}_{q+1}$, $a_{0k} = 9$, $b_{0k} = 4$, for $k = 1, \ldots, r$.
(2) For the prior distributions of $\mathbf{\Delta}, \mathbf{\Delta}_0$ and $\mathbf{\Sigma}_\delta$, the scale matrices $\mathbf{\Psi}_1, \mathbf{\Psi}_2, \mathbf{\Psi}_\delta$ of the Inverse-Wishart distributions are set to 10^{-6} times the identity matrix. The degrees of freedom are set to $v_1 = m, v_2 = p - m$, and $v_0 = q$.
(3) For the prior distributions of $\boldsymbol{c}$ and $\boldsymbol{A}$, the prior covariance matrices are specified as 10^6 times the identity matrix, and the prior means are set to zero.

We use the trace plots of three Markov chains starting from different initial values to check convergence of the algorithm. Figure S1 in the Supplementary Material depicts the trace plots of several randomly selected elements of $\boldsymbol{\beta}$, showing that the Markov chains mixed well within the initial several thousand iterations. Therefore, we run the MCMC algorithm for 16000 iterations, with the first 8000 iterations as the burn-in stage. Four information criteria, AIC, BIC, DIC, and AWE, are employed for model selection. Due to the presence of the latent factor, the complete likelihood can be represented as an integration of $p(\boldsymbol{\theta}, \boldsymbol{\eta} | \mathcal{D})$ with respect to $\boldsymbol{\eta}$, which can be complex. We use the importance sampling approach to approximate the maximized complete data likelihood. In each replication, we calculate the information criteria for different values of m ranging from 0 to p. The case where $m = p$ corresponds to the standard SEM method. The estimated envelope dimension, denoted as $\hat{m}$, is selected based on the minimum criterion value.

The accuracy of the model selection is assessed by determining the percentage of correctly identifying the true envelope dimension m_{true}. The selection rate results for the different settings of m_{true} and p are presented in Table 1. As the sample size n increases, the selected m tends to concentrate more on the true envelope dimension m_{true} for all methods. The selection accuracy varies across different settings of

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Table 1. Percentages of estimated envelope dimension $\hat{m}$ that were selected by the AIC, BIC, AWE, and DIC methods in 100 replications

(a) $m_{\text{true}} = 2,\ p = 20$

	n	$m \le 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m \ge 6$
AIC	50	0.00	0.00	0.00	0.02	0.09	0.89
	150	0.00	0.00	0.03	0.06	0.18	0.73
	300	0.00	0.01	0.04	0.17	0.24	0.54
	600	0.00	0.07	0.11	0.11	0.16	0.55
BIC	50	0.00	0.02	0.02	0.15	0.24	0.57
	150	0.00	0.21	0.19	0.19	0.21	0.20
	300	0.00	0.26	0.21	0.23	0.17	0.13
	600	0.00	0.32	0.28	0.12	0.12	0.16
AWE	50	0.00	0.88	0.10	0.01	0.01	0.00
	150	0.00	0.82	0.17	0.01	0.00	0.00
	300	0.00	0.83	0.15	0.02	0.00	0.00
	600	0.00	0.68	0.25	0.06	0.01	0.00
DIC	50	0.00	0.13	0.12	0.09	0.16	0.50
	150	0.00	0.14	0.07	0.08	0.12	0.59
	300	0.00	0.12	0.12	0.07	0.10	0.59
	600	0.00	0.25	0.21	0.17	0.27	0.10

(b) $m_{\text{true}} = 2,\ p = 40$

	n	$m \le 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m \ge 6$
AIC	50	0.00	0.00	0.00	0.00	0.00	1.00
	150	0.00	0.00	0.00	0.00	0.01	0.99
	300	0.00	0.00	0.00	0.00	0.02	0.98
	600	0.00	0.00	0.00	0.04	0.12	0.84
BIC	50	0.00	0.00	0.00	0.00	0.00	1.00
	150	0.00	0.00	0.00	0.00	0.02	0.98
	300	0.00	0.00	0.01	0.03	0.15	0.81
	600	0.00	0.00	0.08	0.14	0.16	0.62
AWE	50	0.00	0.00	0.00	0.00	0.09	0.91
	150	0.00	0.10	0.24	0.35	0.16	0.15
	300	0.00	0.13	0.26	0.41	0.09	0.11
	600	0.00	0.26	0.20	0.26	0.16	0.12
DIC	50	0.00	0.02	0.07	0.09	0.17	0.65
	150	0.00	0.15	0.11	0.12	0.19	0.43
	300	0.00	0.11	0.21	0.15	0.14	0.39
	600	0.00	0.07	0.17	0.14	0.17	0.45

(c) $m_{\text{true}} = 4,\ p = 20$

	n	$m \le 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m \ge 8$
AIC	50	0.00	0.01	0.02	0.17	0.28	0.52
	150	0.00	0.02	0.03	0.14	0.22	0.59
	300	0.00	0.04	0.03	0.20	0.30	0.43
	600	0.00	0.04	0.10	0.14	0.14	0.58
BIC	50	0.00	0.03	0.23	0.35	0.26	0.13
	150	0.00	0.12	0.37	0.27	0.14	0.10
	300	0.00	0.16	0.34	0.28	0.20	0.02
	600	0.00	0.18	0.34	0.20	0.18	0.10
AWE	50	0.01	0.40	0.54	0.04	0.01	0.00
	150	0.00	0.42	0.56	0.02	0.00	0.00
	300	0.00	0.48	0.47	0.05	0.00	0.00
	600	0.00	0.50	0.45	0.05	0.00	0.00
DIC	50	0.00	0.10	0.08	0.08	0.17	0.57
	150	0.00	0.08	0.09	0.08	0.09	0.66
	300	0.01	0.10	0.08	0.10	0.09	0.62
	600	0.01	0.09	0.14	0.08	0.17	0.51

(d) $m_{\text{true}} = 4,\ p = 40$

	n	$m \le 3$	$m = 4$	$m = 5$	$m = 6$	$m = 7$	$m \ge 8$
AIC	50	0.00	0.00	0.00	0.00	0.00	1.00
	150	0.00	0.00	0.00	0.00	0.01	0.99
	300	0.00	0.00	0.00	0.01	0.06	0.93
	600	0.00	0.00	0.01	0.16	0.12	0.71
BIC	50	0.00	0.00	0.00	0.00	0.00	1.00
	150	0.00	0.00	0.00	0.00	0.06	0.94
	300	0.00	0.00	0.03	0.11	0.20	0.66
	600	0.00	0.01	0.11	0.23	0.17	0.48
AWE	50	0.00	0.00	0.00	0.02	0.12	0.86
	150	0.00	0.07	0.25	0.35	0.16	0.17
	300	0.00	0.11	0.33	0.34	0.11	0.11
	600	0.00	0.18	0.33	0.29	0.11	0.09
DIC	50	0.00	0.05	0.16	0.14	0.12	0.53
	150	0.00	0.13	0.12	0.18	0.15	0.42
	300	0.00	0.13	0.13	0.21	0.11	0.42
	600	0.00	0.11	0.15	0.08	0.14	0.52

(e) $m_{\text{true}} = 6,\ p = 20$

	n	$m \le 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$	$m \ge 10$
AIC	50	0.00	0.06	0.07	0.24	0.32	0.31
	150	0.00	0.07	0.11	0.20	0.26	0.36
	300	0.00	0.16	0.14	0.19	0.18	0.33
	600	0.00	0.13	0.15	0.21	0.15	0.36
BIC	50	0.00	0.36	0.20	0.24	0.16	0.04
	150	0.00	0.41	0.23	0.24	0.07	0.05
	300	0.00	0.40	0.30	0.16	0.11	0.03
	600	0.00	0.28	0.24	0.23	0.10	0.15
AWE	50	0.00	0.97	0.03	0.00	0.00	0.00
	150	0.00	0.96	0.04	0.00	0.00	0.00
	300	0.00	0.91	0.09	0.00	0.00	0.00
	600	0.00	0.67	0.26	0.06	0.00	0.01
DIC	50	0.00	0.10	0.09	0.09	0.12	0.60
	150	0.00	0.07	0.05	0.06	0.16	0.66
	300	0.00	0.10	0.07	0.09	0.08	0.66
	600	0.00	0.08	0.07	0.10	0.13	0.62

(f) $m_{\text{true}} = 6,\ p = 40$

	n	$m \le 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$	$m \ge 10$
AIC	50	0.00	0.00	0.00	0.00	0.00	1.00
	150	0.00	0.00	0.00	0.01	0.01	0.98
	300	0.00	0.00	0.00	0.03	0.10	0.87
	600	0.00	0.00	0.01	0.09	0.13	0.77
BIC	50	0.00	0.00	0.00	0.00	0.01	0.99
	150	0.00	0.00	0.00	0.02	0.09	0.89
	300	0.00	0.00	0.03	0.07	0.20	0.70
	600	0.00	0.01	0.10	0.21	0.24	0.45
AWE	50	0.00	0.00	0.03	0.12	0.25	0.60
	150	0.00	0.12	0.37	0.22	0.21	0.08
	300	0.00	0.18	0.31	0.28	0.14	0.09
	600	0.00	0.24	0.40	0.26	0.08	0.02
DIC	50	0.00	0.05	0.14	0.14	0.16	0.51
	150	0.00	0.08	0.12	0.19	0.24	0.37
	300	0.00	0.09	0.19	0.16	0.20	0.36
	600	0.00	0.12	0.17	0.19	0.17	0.35

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

m_{true}, p, and the true parameter value, and all four methods can exhibit good or poor performance in certain situations. However, among the four methods compared, the AWE method consistently achieves the highest accuracy across all settings. It is important to note that all four methods tend to overestimate the true envelope dimension. This overestimation has also been observed in previous studies (Chakraborty & Su, 2024; Lee et al., 2022). In such cases, the cost of overestimating m may involve losing some efficiency gained from dimension reduction, yet without introducing any bias in parameter estimation, as no material information is lost. The encouraging finding is that despite the fluctuating and dissatisfying selection rate, the overestimated models selected by AWE method still perform comparably to the true envelope model, although there might be a slight loss in efficiency. This assertion is supported by the estimation results of $\boldsymbol{\beta}$, which is one of the main focuses of this analysis.

Figure 1 and Tables 2 and 3 display the root mean squared error (RMSE) and bias of the estimated $\boldsymbol{\beta}$ for two specific settings of m_{true} and p. The plot includes the selected model with $\hat{m}$, the true model with m_{true}, and the standard SEM for two settings of m_{true} and p. The explicit values of the RMSE and bias of several selected elements in $\boldsymbol{\beta}$ are presented in Tables 2 and 3, respectively. The figure and the tables demonstrate that the envelope model can achieve efficiency gains and significantly reduce the RMSE and bias compared to the standard SEM. This improvement is particularly pronounced when the sample size is small. The performance for the other settings is similar, and the figures are not presented. Increasing the number of replications, e.g., to 200, only slightly reduced the RMSE and bias of the $\boldsymbol{\beta}$ estimates further. Figure S2 in the Supplementary Material provides detailed variations observed in the estimated $\boldsymbol{\beta}$ elements under the scenario of $m = 2$, $p = 40$ to illustrate this marginal improvement from additional replications. Moreover, as the sample size increases, the selected model in the envelope model tends to behave similarly to the true envelope model, and their performance in terms of RMSE becomes almost indistinguishable. Regarding the estimation of the latent factor matrix, Table 4 shows that all three methods perform well with small RMSE values, and there is no noticeable difference among them. This suggests that the SEM can incorporate the envelope structure in the structural equation without compromising the accuracy of the measurement equation. The results for the other settings exhibit similar performance and thus are not reported.

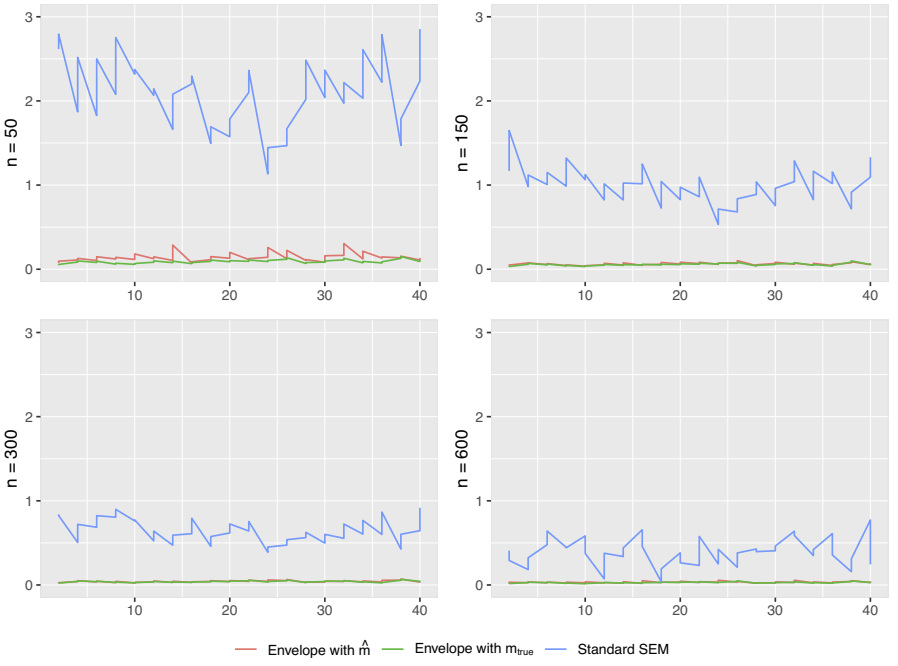
The enhanced estimation accuracy of BESEM gained from the underlying envelope space is further corroborated through a comparison with SEM utilizing BLasso, a widely employed regularized SEM approach (Feng et al., 2015). Table S1 and Figure S3 in the Supplementary Material present the bias and RMSE of the $\boldsymbol{\beta}$ coefficients estimated by the proposed BESEM, standard SEM, and SEM with BLasso. Notably, BESEM exhibits minimal bias and RMSE, standard SEM shows the highest bias and RMSE, whereas SEM with BLasso falls in between.

We conduct several additional simulations to evaluate further the robustness of the proposed method in dimension selection and parameter estimation. The first one considers the scenario, where the lower-dimensional envelope subspace is non-existent, i.e., $m = p$. The second one examines the validity of the proposed method in terms of violating the normality assumptions on the distribution of predictors by generating $\mathbf{X}$ from heavy-tailed and skewed distributions. Additionally, we perform prior sensitivity analyses with respect to different choices of hyperparameters. Performance of the proposed method is stable across the scenarios. Detailed setups and results are provided in the Supplementary Material.

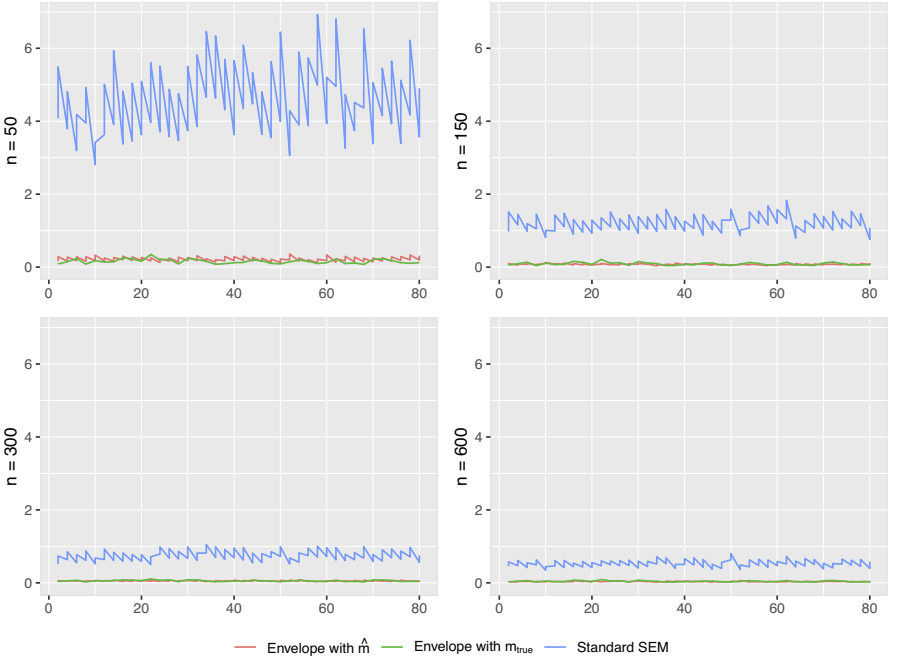
In summary, the envelope model, including the selected model and the true model, outperforms the standard SEM model, especially in estimating $\boldsymbol{\beta}$. While information criteria may exhibit limitations in terms of selection rate, it is noteworthy that as the sample size increases, the selected model within the envelope model framework can approximate the true model closely. This observation highlights the feasibility of the AWE method in selecting appropriate envelope models and maintaining their effectiveness in estimating the parameter of interest efficiently.

To obtain the results reported in Table 1 using the proposed method, the computational time per replication ranges from about 1 min to a maximum of approximately 17 min on a Linux machine running R with a CPU block speed of 2.60 GHz. The actual timing depends on the specific combination of sample size, predictor dimension, and envelope dimension. Detailed results are summarized in

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press



(a) $m_{\text{true}} = 2$ and $p = 20$.



(b) $m_{\text{true}} = 2$ and $p = 40$.

Figure 1. The RMSE of the estimated elements of $\boldsymbol{\beta}$ for two cases in simulation.
Note: x-axis: coordinate of $\boldsymbol{\beta}$. y-axis: value of RMSE.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Table 2. The RMSE of estimated elements in $\boldsymbol{\beta}$ for two cases in Simulation

(a) $m_{\text{true}} = 2$, $p = 20$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
$\beta_{1,1}$	0.0952	0.0494	0.0269	0.0226	0.0568	0.0326	0.0224	0.0166	2.7929	1.6482	0.8245	0.2910
$\beta_{13,1}$	0.2243	0.1023	0.0650	0.0419	0.1361	0.0817	0.0591	0.0492	1.6694	0.8369	0.5374	0.3805
$\beta_{7,2}$	0.1040	0.0497	0.0309	0.0212	0.0777	0.0460	0.0312	0.0217	1.6624	0.8231	0.4722	0.3380
$\beta_{15,2}$	0.0826	0.0685	0.0345	0.0256	0.0843	0.0558	0.0403	0.0241	2.0392	0.7557	0.4977	0.4065

(b) $m_{\text{true}} = 2$, $p = 40$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
$\beta_{5,1}$	0.3311	0.1278	0.0639	0.0419	0.1793	0.1117	0.0609	0.0514	3.4134	1.0059	0.6798	0.4462
$\beta_{18,1}$	0.2064	0.0805	0.0616	0.0483	0.0786	0.0426	0.0337	0.0195	6.3355	1.5865	0.9877	0.6866
$\beta_{30,1}$	0.3365	0.0681	0.0537	0.0319	0.1210	0.0676	0.0449	0.0384	5.1997	1.5765	0.9637	0.5868
$\beta_{1,2}$	0.1630	0.0610	0.0363	0.0236	0.0832	0.0537	0.0362	0.0262	4.0754	0.9764	0.5194	0.4569
$\beta_{25,2}$	0.1244	0.0436	0.0305	0.0249	0.0892	0.0480	0.0366	0.0256	3.9930	1.2857	0.7239	0.6305
$\beta_{32,2}$	0.1735	0.0594	0.0353	0.0309	0.0964	0.0612	0.0352	0.0280	3.2580	0.7964	0.6118	0.4324

Note: Displayed are several randomly selected elements in $\boldsymbol{\beta}$.

Table 3. The bias of estimated elements in $\boldsymbol{\beta}$ for two cases in simulation

(a) $m_{\text{true}} = 2$, $p = 20$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
$\beta_{1,1}$	−0.0038	0.0025	−0.0025	0.0047	−0.0084	−0.0073	−0.0004	−0.0035	−0.0069	0.0887	−0.0571	−0.2859
$\beta_{13,1}$	0.0196	0.0035	−0.0101	−0.0001	−0.0083	0.0095	0.0004	0.0045	−0.1245	0.1159	−0.0516	−0.0472
$\beta_{7,2}$	−0.0054	0.0000	−0.0036	−0.0022	−0.0022	−0.0087	−0.0038	−0.0029	−0.0516	−0.0532	0.0351	−0.3362
$\beta_{15,2}$	−0.0045	−0.0041	0.0058	−0.0005	−0.0006	0.0032	0.0024	0.0018	0.2290	0.0346	−0.0659	−0.4060

(b) $m_{\text{true}} = 2$, $p = 40$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
$\beta_{5,1}$	−0.0303	0.0201	−0.0004	0.0001	0.0294	0.0092	0.0033	0.0111	−0.0807	0.1157	0.0380	−0.0399
$\beta_{18,1}$	0.0325	−0.0121	0.0013	0.0121	−0.0181	−0.0070	−0.0096	0.0031	0.5396	0.2090	−0.0989	−0.0599
$\beta_{30,1}$	0.0121	0.0005	−0.0058	−0.0062	−0.0355	−0.0070	−0.0030	−0.0104	−0.4623	0.0628	−0.1453	−0.0381
$\beta_{1,2}$	−0.0114	0.0103	−0.0073	−0.0009	−0.0105	0.0008	−0.0035	0.0001	0.1144	−0.1341	0.0061	0.0582
$\beta_{25,2}$	−0.0024	−0.0065	−0.0014	−0.0044	−0.0146	−0.0002	−0.0053	−0.0037	0.0876	−0.3007	−0.0536	0.0771
$\beta_{32,2}$	−0.0067	−0.0028	−0.0021	0.0029	−0.0250	0.0054	−0.0007	−0.0025	−0.2587	0.0071	−0.1480	−0.0596

Note: Displayed are several randomly selected elements in $\boldsymbol{\beta}$.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Table 4. The RMSE of estimated free elements in $\mathbf{\Lambda}$ for two cases in simulation

(a) $m_{\text{true}} = 2$, $p = 20$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
λ_1	0.0014	0.0008	0.0006	0.0003	0.0015	0.0007	0.0006	0.0004	0.0013	0.0007	0.0006	0.0008
λ_2	0.0038	0.0019	0.0015	0.0011	0.0039	0.0021	0.0015	0.0010	0.0038	0.0021	0.0015	0.0014
λ_3	0.0044	0.0030	0.0017	0.0014	0.0047	0.0028	0.0019	0.0014	0.0051	0.0028	0.0018	0.0012
λ_4	0.0041	0.0025	0.0015	0.0013	0.0044	0.0023	0.0017	0.0013	0.0047	0.0022	0.0015	0.0007

(b) $m_{\text{true}} = 2$, $p = 40$

Method	Envelope $\hat{m}$				Envelope m_{true}				SEM			
n	50	150	300	600	50	150	300	600	50	150	300	600
λ_1	0.0027	0.0014	0.0011	0.0008	0.0029	0.0016	0.0011	0.0007	0.0028	0.0017	0.0009	0.0007
λ_2	0.0027	0.0014	0.0012	0.0008	0.0030	0.0016	0.0011	0.0007	0.0027	0.0017	0.0010	0.0007
λ_3	0.0021	0.0012	0.0009	0.0006	0.0019	0.0012	0.0008	0.0006	0.0021	0.0011	0.0007	0.0006
λ_4	0.0047	0.0023	0.0017	0.0011	0.0042	0.0024	0.0017	0.0010	0.0043	0.0023	0.0018	0.0010

Table S2 in the Supplementary Material. The code for implementing the preceding analysis is scripted in R with Rcpp and is freely available in the Supplementary Material.

5. Real data analysis

We apply the proposed BESEM to analyze the ADNI dataset. The ADNI project, launched in 2004, aims to advance research on the early detection, diagnosis, tracking, and treatment of AD and other forms of cognitive impairment. It has become a widely recognized and extensively studied collection of neuroimaging, clinical, and biomarker data related to AD. One crucial topic in the study of AD and other dementia diseases is understanding the relationship between cognitive decline and changes in brain structure. Previous studies have indicated that various brain regions are affected to different extents in AD (Bartos et al., 2019; Jones et al., 2006). In our study, we integrate neuroimaging phenotypes comprising volumetric and cortical thickness measures derived from MRI data. These measures capture the structural characteristics of the brain across a moderate number of ROIs. We incorporate these neuroimaging phenotypes into a regression model and utilize the envelope technique to gain new insights into the relationship between brain structure and cognitive ability.

Clinical diagnostic tools, including neurological exams and cognitive and functional assessments, are believed to reflect the degree of cognitive impairment and monitor disease progression (Albert et al., 2013). Therefore, we regard cognitive impairment as a latent variable denoted by η. We consider five different test scores from the ADNI study to assess cognitive impairment in various domains. The first test score is the Mini-Mental State Examination (MMSE), a widely used screening tool that assesses cognitive impairment from various cognitive domains, including orientation, attention and calculation, recall, language, and visuospatial abilities. The second test score is the AD Assessment Scale-Cognitive subscale 13 (ADAS13), which is specifically designed for use in clinical trials to measure cognitive changes associated with AD. The scoring of ADAS13 is more complex than that of MMSE. Higher ADAS13 scores indicate severe cognitive impairment. The third one is the Functional Assessment Questionnaire (FAQ), which evaluates an individual's functional ability to perform activities of daily living such as dressing, bathing, grooming, eating, and managing finances. The final two test scores are

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

derived from the Rey Auditory Verbal Learning Test (RAVLT), a neuropsychological test that assesses verbal learning and memory. Specifically, we selected two measures from RAVLT: RAVLT Immediate (RAVLT.immediate) and RAVLT percent forgetting (RAVLT.perc.forgetting). These two measures are selected since they capture different aspects of episodic memory and have been shown to be closely related to AD detection by previous research (Estévez-González et al., 2003; Moradi et al., 2017). To ensure consistency in interpreting the test scores, we reversed the MMSE and RAVLT.immediate scores so that higher values indicate a severe impairment in cognitive function. Additionally, all five test scores were standardized before conducting the analysis. These standardized test scores are denoted as $\mathbf{Y} = (Y_1, \ldots, Y_5)^T$, as the response variables in equation (8).

The MRI brain imaging data were preprocessed by UCSF using Freesurfer methods, which involved segmenting the T1 weighted images into small regions to define volumetric and cortical thickness values. Detailed parcellation and quality control (QC) guidelines can be found on the LONI-ADNI website. We followed the instructions provided by Szefer et al. (2017), and Greenlaw et al. (2017) and include 56 derived neroimaging phenotypes. A description of these phenotypes can be found in Table 5. The log-transformed values of the volumetric or cortical thickness measures serve as the predictors $\mathbf{X} = (X_1, \ldots, X_{56})^T$ in the structural envelope model.

The original dataset from the ADNI-1 phase comprises clinical and imaging measurements from 800 participants, with missing data in both cognitive test scores and brain imaging data. After excluding missing data, a total of $n = 557$ observations are retained in the analysis. Among these observations, 169 individuals are diagnosed with cognitive normal (CN), 271 have mild cognitive impairment (MCI), and 117 are diagnosed with AD. We apply the proposed BESEM method to this dataset, using a flat prior similar to that described in Section 4.

The MCMC algorithm converges within the initial thousands of iterations, as shown by the trace plots depicted in Figure S4 in the Supplementary Material. Therefore, a total of 20000 samples are drawn from the conditional posterior density, with the first 8000 iterations discarded as the burn-in stage. The AIC, BIC, DIC, and AWE methods were used to determine the optimal dimension of the envelope structure. All these methods consistently indicate that the optimal dimension is $m = 1$. Therefore, the subsequent estimation results using the BESEM method are based on the envelope model with a dimension of 1.

The point estimate of the factor loading $\boldsymbol{\Lambda}$ is $(1.000, 0.618, 0.787, 0.914, 0.730)^T$, with a negligible standard error estimate for each element. The estimates of coefficients $\boldsymbol{\beta}$ (Est), along with their standard error estimates (SE), and 95% credible interval (CI) of the posterior samples, are displayed in Figure 2a and Table 6. Notably, the inferior lateral ventricle and lateral ventricle for both hemispheres are found to have a significant effect under the envelope structure, while the remaining ROIs exhibit minimal impact, with estimated coefficients close to zero. This finding aligns with previous research by Bartos et al. (2019), which suggests that the inferior regions of both lateral ventricles have the most significant enlargements in individuals with AD.

For comparison, we also present the results obtained using the standard SEM method in Figure 2b and Table 7. Akin to the simulation study, the standard SEM was estimated using a Bayesian approach on the subset of 557 subjects with complete measurements, ensuring a fair comparison with the proposed BESEM. In line with standard Bayesian SEM practices, we assigned conjugate normal-inverse-gamma priors to parameters in the measurement equation and vague normal-inverse-Wishart priors to parameters in the structural equation to derive Bayesian estimates. These results demonstrate a conspicuous deviation from those obtained using the envelope method. In this analysis, we exclude regions with point estimates close to zero and those with 95% CIs that include zero and identify several ROIs that exhibited significance in the regression model. The selected regions are marked with bold text in Table 7. The selected significant regions do not exhibit consistency between the two brain hemispheres. The fusiform and inferior temporal in the left hemisphere is shown to be more significant, which aligns with the notion proposed by Galton et al. (2001). Meanwhile, we also found a significant impact on the hippocampus volume in the left hemisphere but not in the right hemisphere. This finding contradicts the conclusion in Barnes et al. (2005) and Yang et al. (2017), which suggests a preferential association of the right hippocampus with AD. Such partial inconsistency with the existing literature

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

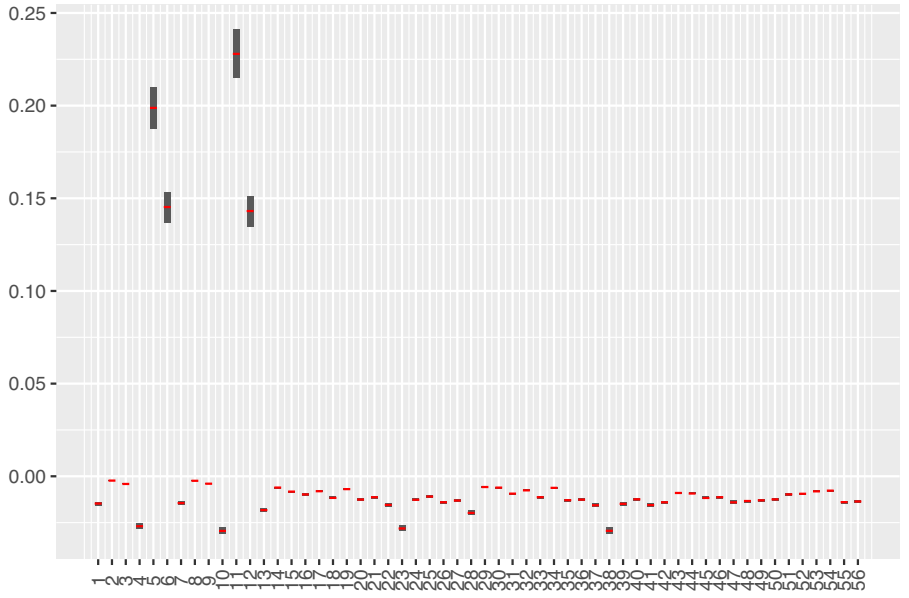
Table 5. Imaging phenotypes defined as volumetric or cortical thickness measures of $28 \times 2 = 56$ ROIs

ID	Measurement	ROI
AmygVol	Volume	Amygdala
CerebCtx	Volume	Cerebral cortex
CerebWM	Volume	Cerebral white matter
HippVol	Volume	Hippocampus
InfLatVent	Volume	Inferior lateral ventricle
LatVent	Volume	Lateral ventricle
EntCtx	Thickness	Entorhinal cortex
Fusiform	Thickness	Fusiform gyrus
InfParietal	Thickness	Inferior parietal gyrus
InfTemporal	Thickness	Inferior temporal gyrus
MidTemporal	Thickness	Middle temporal gyrus
Parahipp	Thickness	Parahippocampal gyrus
PostCing	Thickness	Posterior cingulate
Postcentral	Thickness	Postcentral gyrus
Precentral	Thickness	Precentral gyurs
Precuneus	Thickness	Precuneus
SupFrontal	Thickness	Superior frontal gyrus
SupParietal	Thickness	Superior parietal gyrus
SupTemporal	Thickness	Superior temporal gyrus
Supramarg	Thickness	Supramarginal gyrus
TemporalPole	Thickness	Temporal pole
MeanCing	Mean thickness	Caudal anterior cingulate, isthmus cingulate, posterior cingulate, rostral anterior cingulate
MeanFront	Mean thickness	Caudal midfrontal, rostral midfrontal, superior frontal, lateral orbitofrontal, and medial orbitofrontal gyri, frontal pole
MeanLatTemp	Mean thickness	Inferior temporal, middle temporal, and superior temporal gyri
MeanMedTemp	Mean thickness	Fusiform, parahippocampal, and lingual gyri, temporal pole and transverse temporal pole
MeanPar	Mean thickness	Inferior and superior parietal gyri, supramarginal gyrus, and precuneus
MeanSensMotor	Mean thickness	Precentral and postcentral gyri
MeanTemp	Mean thickness	Inferior temporal, middle temporal, superior temporal, fusiform, parahippocampal, lingual gyri, temporal pole, transverse temporal pole

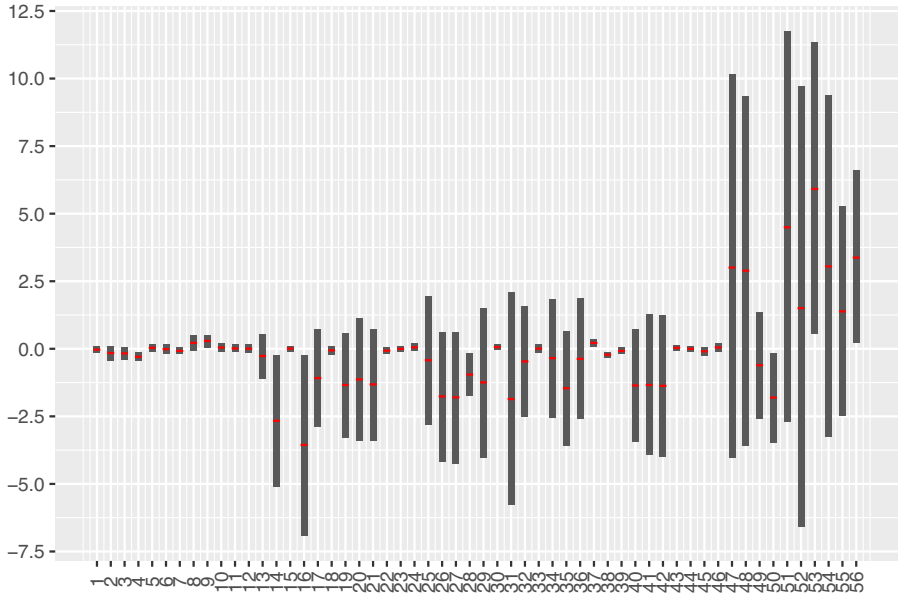
Note: Each phenotype in the table corresponds to the two phenotypes for the left and right hemispheres.

is likely a consequence of multicollinearity among the covariates and overfitting to the noise or the immaterial information in the multivariate regression problem, thereby raises doubts about the validity of the results. It demonstrates that the standard SEM model may yield results lacking interpretability or mask some significant associations in the setting of complex covariates that can be projected to a reducing subspace. The estimation results of SEM with BLasso are also presented in Table S3 and Figure S5 in the Supplementary Material for a comprehensive comparison. A similar pattern to the standard SEM is observed. Moreover, in this dataset, the changes observed in different ROIs are not entirely

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press



(a) BESEM



(b) Standard SEM

Figure 2. Point and 95% interval estimates of each element of $\boldsymbol{\beta}$ for the ADNI study.
Note: x-axis: ID of each ROI, which aligns with the order in Table 5 (adjacent numbers represent the left hemisphere and right hemisphere, respectively). y-axis: Estimated value. Red short line: the value of the estimated coefficient. Grey rectangle: 95% CI.

independent, and a plain linear regression model may not adequately capture the interdependencies among the covariates in such cases, further contributing to noisy and potentially misleading results. In contrast, the BESEM method enhances the identification of important predictors and significantly improves estimation efficiency. It better handles the challenges posed by high-dimensional covariates and accounts for the interdependencies among ROIs, leading to more reliable and informative results.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Table 6. Point estimates (Est), Standard Error Estimates (SE), and 95% CIs of the elements in $\boldsymbol{\beta}$ by BESEM method in ADNI study

(a) Left hemisphere

ID	Est	SE	95% CI
AmygVol	−0.015	0.001	(−0.0172, −0.0123)
CerebCtx	−0.002	0.000	(−0.0027, −0.0019)
CerebWM	−0.004	0.000	(−0.0048, −0.0035)
HippVol	−0.027	0.002	(−0.0315, −0.0225)
InfLatVent	0.199	0.017	(0.1657, 0.2322)
LatVent	0.145	0.012	(0.1211, 0.1697)
EntCtx	−0.018	0.002	(−0.0213, −0.0152)
Fusiform	−0.006	0.001	(−0.0071, −0.0051)
InfParietal	−0.008	0.001	(−0.0097, −0.0069)
InfTemporal	−0.010	0.001	(−0.0115, −0.0082)
MidTemporal	−0.008	0.001	(−0.0094, −0.0067)
Parahipp	−0.012	0.001	(−0.0136, −0.0097)
PostCing	−0.007	0.001	(−0.0081, −0.0058)
Postcentral	−0.013	0.001	(−0.0148, −0.0103)
Precentral	−0.011	0.001	(−0.0132, −0.0094)
Precuneus	−0.015	0.001	(−0.0181, −0.0129)
SupFrontal	−0.028	0.002	(−0.0329, −0.0235)
SupParietal	−0.013	0.001	(−0.0148, −0.0105)
SupTemporal	−0.011	0.001	(−0.0128, −0.0091)
Supramarg	−0.014	0.001	(−0.0164, −0.0117)
TemporalPole	−0.013	0.001	(−0.0153, −0.0109)
MeanCing	−0.009	0.001	(−0.0105, −0.0075)
MeanFront	−0.009	0.001	(−0.0108, −0.0077)
MeanLatTemp	−0.012	0.001	(−0.0137, −0.0097)
MeanMedTemp	−0.011	0.001	(−0.0132, −0.0094)
MeanPar	−0.014	0.001	(−0.0164, −0.0117)
MeanSensMotor	−0.013	0.001	(−0.0156, −0.0111)
MeanTemp	−0.013	0.001	(−0.0152, −0.0108)

(b) Right hemisphere

ID	Est	SE	95% CI
AmygVol	−0.014	0.001	(−0.0169, −0.0121)
CerebCtx	−0.002	0.000	(−0.0028, −0.0020)
CerebWM	−0.004	0.000	(−0.0047, −0.0033)
HippVol	−0.029	0.002	(−0.0344, −0.0246)
InfLatVent	0.228	0.019	(0.1900, 0.2662)
LatVent	0.143	0.012	(0.1193, 0.1672)
EntCtx	−0.020	0.002	(−0.0231, −0.0165)
Fusiform	−0.006	0.000	(−0.0068, −0.0048)
InfParietal	−0.006	0.001	(−0.0073, −0.0050)
InfTemporal	−0.009	0.001	(−0.0109, −0.0078)
MidTemporal	−0.008	0.001	(−0.0089, −0.0063)
Parahipp	−0.011	0.001	(−0.0132, −0.0094)
PostCing	−0.006	0.001	(−0.0073, −0.0052)
Postcentral	−0.013	0.001	(−0.0151, −0.0108)
Precentral	−0.013	0.001	(−0.0147, −0.0105)
Precuneus	−0.016	0.001	(−0.0183, −0.0130)
SupFrontal	−0.029	0.002	(−0.0344, −0.0245)
SupParietal	−0.015	0.001	(−0.0174, −0.0124)
SupTemporal	−0.012	0.001	(−0.0145, −0.0104)
Supramarg	−0.016	0.001	(−0.0182, −0.0130)
TemporalPole	−0.014	0.001	(−0.0164, −0.0117)
MeanCing	−0.012	0.001	(−0.0146, −0.0104)
MeanFront	−0.010	0.001	(−0.0114, −0.0081)
MeanLatTemp	−0.009	0.001	(−0.0110, −0.0079)
MeanMedTemp	−0.008	0.001	(−0.0095, −0.0068)
MeanPar	−0.008	0.001	(−0.0090, −0.0065)
MeanSensMotor	−0.014	0.001	(−0.0163, −0.0116)
MeanTemp	−0.014	0.001	(−0.0159, −0.0113)

6. Conclusion and discussion

This article introduces a BESEM that incorporates a predictor envelope structure in the structural equation, in conjunction with a factor model, to achieve dimension reduction. The response variables are grouped into latent factors with a predetermined structure, allowing for dimension reduction of multiple responses in the factor model. In the structural equation, the concerned predictors are decomposed into two parts, the material part and the immaterial part. These two parts are independent, with the immaterial part containing no information relevant to the latent factor. The presence of the envelope structure, along with a small envelope dimension, can greatly reduce the number of unknown

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Table 7. Point estimates (Est), Standard Error Estimates (SE), and 95% CI of the elements in $\boldsymbol{\beta}$ by standard SEM method in ADNI study

(a) Left hemisphere

ID	Est	SE	95% CI
AmygVol	−0.030	0.061	(−0.1493, 0.0893)
CerebCtx	−0.156	0.140	(−0.4363, 0.1105)
CerebWM	−0.172	0.118	(−0.4033, 0.0584)
HippVol	−0.296	0.080	(−0.4514, −0.1396)
InfLatVent	0.035	0.072	(−0.1083, 0.1731)
LatVent	−0.013	0.087	(−0.1829, 0.1557)
EntCtx	−0.268	0.414	(−1.0912, 0.5420)
Fusiform	−2.665	1.241	(−5.1143, −0.2328)
InfParietal	0.008	0.049	(−0.0868, 0.1053)
InfTemporal	−3.557	1.709	(−6.9172, −0.2254)
MidTemporal	−1.083	0.917	(−2.8919, 0.7291)
Parahipp	−0.059	0.084	(−0.2232, 0.1042)
PostCing	−1.340	0.988	(−3.2790, 0.5975)
Postcentral	−1.134	1.158	(−3.3929, 1.1335)
Precentral	−1.318	1.042	(−3.3820, 0.7286)
Precuneus	−0.068	0.066	(−0.1972, 0.0580)
SupFrontal	−0.004	0.052	(−0.1066, 0.0968)
SupParietal	0.054	0.069	(−0.0820, 0.1893)
SupTemporal	−0.422	1.198	(−2.8027, 1.9450)
Supramarg	−1.762	1.225	(−4.1690, 0.6299)
TemporalPole	−1.796	1.236	(−4.2305, 0.6222)
MeanCing	0.035	0.050	(−0.0652, 0.1316)
MeanFront	−0.001	0.050	(−0.1000, 0.0979)
MeanLatTemp	−0.094	0.082	(−0.2560, 0.0664)
MeanMedTemp	0.049	0.079	(−0.1018, 0.2038)
MeanPar	3.006	3.661	(−4.0389, 10.1664)
MeanSensMotor	2.890	3.304	(−3.5973, 9.3247)
MeanTemp	−0.606	0.996	(−2.5886, 1.3453)

(b) Right hemisphere

ID	Est	SE	95% CI
AmygVol	−0.077	0.062	(−0.1971, 0.0430)
CerebCtx	0.217	0.142	(−0.0592, 0.5002)
CerebWM	0.291	0.121	(0.0525, 0.5241)
HippVol	0.046	0.079	(−0.1098, 0.2016)
InfLatVent	0.014	0.066	(−0.1155, 0.1449)
LatVent	0.006	0.081	(−0.1532, 0.1638)
EntCtx	−0.952	0.392	(−1.7415, −0.1869)
Fusiform	−1.243	1.412	(−4.0161, 1.5097)
InfParietal	0.062	0.048	(−0.0317, 0.1538)
InfTemporal	−1.858	2.015	(−5.7815, 2.0930)
MidTemporal	−0.468	1.034	(−2.5246, 1.5601)
Parahipp	0.006	0.084	(−0.1557, 0.1705)
PostCing	−0.344	1.101	(−2.5441, 1.8310)
Postcentral	−1.460	1.086	(−3.5695, 0.6496)
Precentral	−0.373	1.127	(−2.5993, 1.8520)
Precuneus	0.216	0.069	(0.0829, 0.3505)
SupFrontal	−0.218	0.052	(−0.3199, −0.1172)
SupParietal	−0.079	0.062	(−0.1996, 0.0405)
SupTemporal	−1.359	1.060	(−3.4502, 0.7165)
Supramarg	−1.340	1.320	(−3.9382, 1.2594)
TemporalPole	−1.373	1.339	(−4.0058, 1.2367)
MeanCing	−1.805	0.834	(−3.4736, −0.1749)
MeanFront	4.497	3.663	(−2.7019, 11.7353)
MeanLatTemp	1.502	4.105	(−6.6029, 9.7027)
MeanMedTemp	5.914	2.759	(0.5552, 11.3317)
MeanPar	3.047	3.229	(−3.2699, 9.3676)
MeanSensMotor	1.386	1.967	(−2.4794, 5.2762)
MeanTemp	3.377	1.626	(0.2048, 6.5945)

parameters, leading to improved estimation efficiency. The application of the proposed model to the ADNI dataset demonstrates its potential in identifying important covariates.

While there are still several areas of interest that warrant further investigation. First, there is a need to develop an efficient method for selecting the envelope dimension. Currently, we approach this as a model selection task and utilize information criteria to identify the optimal envelope structure, which can lead to overestimated envelope dimension. A potential solution to mitigate this limitation involves circumventing the selection of envelope dimension by developing a weighted average variant of the

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

envelope estimator, using the information criterion values as potential weights (Eck & Cook, 2017). Nonetheless, this approach may introduce theoretical and computational challenges, and we consider it as a promising future direction that warrants continuing effort. Besides, employing information criteria for envelope dimension can be time-consuming and computationally intensive. To address this, Zhang & Mai (2018) proposed two unified approaches, full Grassmannian (FG) optimization and 1 dimension (1D) selection, which provide theoretical justifications for selection consistency and exhibit computational stability. Adapting this idea from a Bayesian perspective holds promise, and we aim to develop a method that can simultaneously select the dimension and perform estimation. Second, the substitution of the CFA model with an EFA model is worth exploring. This would relax the assumption of a predetermined number of latent factors and instead allow the factor loading matrix to be determined by both prior information and data. However, determining the number of latent factors can be challenging, especially when it is linked to predictors with an uncertain envelope structure. Third, intricate missing mechanisms can arise in multivariate regression while addressing missing data within envelope models is an ongoing area of research with few established methodologies. We see the accommodation of diverse missing mechanisms in the joint modeling of envelope and SEM as a potential for further exploration. Lastly, it is worth considering the possibility of incorporating the envelope structure into the multivariate variables (responses) in the measurement component of an SEM. Cook & Zhang (2015b) introduced envelopes that simultaneously reduce predictors and responses in a multivariate linear regression model, leading to significant improvements over traditional methods. Therefore, we consider it a promising future direction to further extend the BESEM to incorporate an envelope structure for the multivariate responses in the measurement component. For instance, if we can appropriately separate the immaterial part of the response variables in the measurement equation and group the material part into the latent factors, then it is possible to handle high-dimensional response vectors in BESEM. Substantial efforts are required to achieve this advancement.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2025.10027.

Author contributions. C. W. and R. S. are co-first authors. They contributed equally to this work.

Funding statement. Dr. Feng's research was fully supported by the National Natural Science Foundation of China (No. 72271060). Dr. Song's research was fully supported by GRF Grant (No.14303622) from the Research Grant Council of the Hong Kong Special Administration Region.

Competing interests. The authors declare none.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2013). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Focus*, *11*(1), 96–106.

Barnes, J., Scahill, R. I., Schott, J. M., Frost, C., Rossor, M. N., & Fox, N. C. (2005). Does Alzheimer's disease affect hippocampal asymmetry? Evidence from a cross-sectional and longitudinal volumetric MRI study. *Dementia and Geriatric Cognitive Disorders*, *19*(5–6), 338–344.

Bartos, A., Gregus, D., Ibrahim, I., & Tintěra, J. (2019). Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using Freesurfer 6.0. *Psychiatry Research: Neuroimaging*, *287*, 70–74.

Chakraborty, S., & Su, Z. (2024). A comprehensive Bayesian framework for envelope models. *Journal of the American Statistical Association*, *119*(547), 2129–2139.

Chen, T., Su, Z. H., Yang, Y., & Ding, S. S. (2020). Efficient estimation in expectile regression using envelope models. *Electronic Journal of Statistics*, *14*(1), 143–173.

Cinlar, E. (2013). *Introduction to stochastic processes*. Courier Corporation.

Cook, R. D., Forzani, L., & Su, Z. H. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis*, *150*, 42–54.

Cook, R. D., Helland, I., & Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *75*(5), 851–877.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Cook, R. D., Li, B., & Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, *20*(3), 927–960.

Cook, R. D., & Zhang, X. (2015a). Foundations for envelope models and methods. *Journal of the American Statistical Association*, *110*(510), 599–611.

Cook, R. D., & Zhang, X. (2015b). Simultaneous envelopes for multivariate linear regression. *Technometrics*, *57*(1), 11–25.

Ding, S., & Cook, R. D. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *80*(2), 387–408.

Eck, D. J., & Cook, R. D. (2017). Weighted envelope estimation to handle variability in model selection. *Biometrika*, *104*(3), 743–749.

Estévez-González, A., Kulisevsky, J., Boltes, A., Otermín, P., & García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: comparison with mild cognitive impairment and normal aging. *International Journal of Geriatric Psychiatry*, *18*(11), 1021–1028.

Feng, X., Wang, G., Wang, Y., & Song, X. (2015). Structure detection of semiparametric structural equation models with Bayesian adaptive group lasso. *Statistics in Medicine*, *34*(9), 1527–1547.

Feng, X. N., Wang, Y., Lu, B., & Song, X. Y. (2017). Bayesian regularized quantile structural equation models. *Journal of Multivariate Analysis*, *154*, 234–248.

Galton, C. J., Patterson, K., Graham, K., Lambon-Ralph, M. A., Williams, G., Antoun, N., Sahakian, B., & Hodges, J. (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology*, *57*(2), 216–225.

Greenlaw, K., Szefer, E., Graham, J., Lesperance, M., Nathoo, F. S., & Alzheimer's Disease Neuroimaging Initiative. (2017). A Bayesian group sparse multi-task regression model for imaging genetics. *Bioinformatics*, *33*(16), 2513–2522.

Jones, B. F., Barnes, J., Uylings, H. B., Fox, N. C., Frost, C., Witter, M. P., & Scheltens, P. (2006). Differential regional atrophy of the cingulate gyrus in Alzheimer disease: A volumetric MRI study. *Cerebral Cortex*, *16*(12), 1701–1708.

Khare, K., Pal, S., & Su, Z. H. (2017). A Bayesian approach for envelope models. *The Annals of Statistics*, *45*(1), 196–222.

Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, *81*(11), 1565–1578.

Lee, M., Chakraborty, S., & Su, Z. H. (2022). A Bayesian approach to envelope quantile regression. *Statistica Sinica*, *32*, 2339–2357.

Lee, S. Y., & Song, X. Y. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, *60*(3), 624–636.

Moradi, E., Hallikainen, I., Hänninen, T., Tohka, J., & Alzheimer's Disease Neuroimaging Initiative. (2017). Rey's auditory verbal learning test scores can be predicted from whole brain MRI in Alzheimer's disease. *NeuroImage: Clinical*, *13*, 415–427.

Pan, D., Kang, K., Wang, C. J., & Song, X. Y. (2019). Bayesian proportional hazards model with latent variables. *Statistical Methods in Medical Research*, *28*(4), 986–1002.

Rekabdarkolaee, H. M., Wang, Q., Naji, Z., & Fuente, M. (2020). New parsimonious multivariate spatial model: Spatial envelope. *Statistica Sinica*, *30*(3), 1583–1604.

Roy, J., & Lin, X. H. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*, *56*(4), 1047–1054.

Shen, Y. B., Park, Y. H., Chakraborty, S., & Zhang, C. M. (2023). Bayesian simultaneous partial envelope model with application to an imaging genetics analysis. *The New England Journal of Statistics in Data Science*, *1*(2), 237–269.

Song, X. Y., & Lee, S. Y. (2006). A maximum likelihood approach for multisample nonlinear structural equation models with missing continuous and dichotomous data. *Structural Equation Modeling*, *13*(3), 325–351.

Song, X. Y., & Lee, S. Y. (2012). *Basic and advanced Bayesian structural equation modeling: With applications in the medical and behavioral sciences*. John Wiley & Sons.

Song, X. Y., Lu, Z. H., Hser, Y. I., & Lee, S. Y. (2011). A Bayesian approach for analyzing longitudinal structural equation models. *Structural Equation Modeling*, *18*(2), 183–194.

Su, Z., Zhu, G., Chen, X., & Yang, Y. (2016). Sparse envelope model: efficient estimation and response variable selection in multivariate linear regression. *Biometrika*, *103*(3), 579–593.

Su, Z. H., & Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika*, *98*(1), 133–146.

Szefer, E., Lu, D., Nathoo, F., Beg, M. F., Graham, J., & Alzheimer's Disease Neuroimaging Initiative (2017). Multivariate association between single-nucleotide polymorphisms in Alzgene linkage regions and structural changes in the brain: discovery, refinement and validation. *Statistical Applications in Genetics and Molecular Biology*, *16*(5–6), 367–386.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, *82*(398), 528–540.

Wang, X. Q., Song, X. Y., & Zhu, H. T. (2021). Bayesian latent factor on image regression with nonignorable missing data. *Statistics in Medicine*, *40*(4), 920–932.

Wang, Y., Feng, X. N., & Song, X. Y. (2016). Bayesian quantile structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(2), 246–258.

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press

Yang, C., Zhong, S., Zhou, X., Wei, L., Wang, L., & Nie, S. (2017). The abnormality of topological asymmetry between hemispheric brain white matter networks in Alzheimer's disease and mild cognitive impairment. *Frontiers in Aging Neuroscience*, *9*, 261.

Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437–447.

Zhang, X., & Mai, Q. (2018). Model-free envelope dimension selection. *Electronic Journal of Statistics*, *12*(2), 2193–2216.

Cite this article: Wang, C., Sun, R., Feng, X. and Song, X., (2025). Bayesian Structural Equation Envelope Model. *Psychometrika*, 1236–1257. https://doi.org/10.1017/psy.2025.10027

https://doi.org/10.1017/psy.2025.10027 Published online by Cambridge University Press