This is a "preproof" accepted article for *Psychometrika*.
This version may be subject to change during the production process.
DOI: 10.1017/psy.2025.7

Generalized Latent Variable Models for Location, Scale, and Shape parameters

Camilo A. Cárdenas-Hurtado* Irini Moustaki†
Yunxiao Chen‡ Giampiero Marra§

Abstract

We introduce a general framework for latent variable modeling, named Generalized Latent Variable Models for Location, Scale, and Shape parameters (GLVM-LSS). This framework extends the generalized linear latent variable model beyond the exponential family distributional assumption and enables the modeling of distributional parameters other than the mean (location parameter), such as scale and shape parameters, as functions of latent variables. Model parameters are estimated via maximum likelihood. We present two real-world applications on public opinion research and educational testing, and evaluate the model's performance in terms of parameter recovery through extensive simulation studies. Our results suggest that the GLVM-LSS is a valuable tool in applications where modeling higher-order moments of the observed variables through latent variables is of substantive interest. The proposed model is implemented in the `R` package `glvmlss`, available online.

Key words: Latent variable models, Distributional regression, GAMLSS, EM algorithm, Heteroscedasticity, Skewness.

*Corresponding author. Department of Statistics, LSE, c.a.cardenas-hurtado@lse.ac.uk
†Department of Statistics, LSE, i.moustaki@lse.ac.uk
‡Department of Statistics, LSE, y.chen186@lse.ac.uk
§Department of Statistical Science, UCL, giampiero.marra@ucl.ac.uk

1. Introduction

Latent variable models (LVMs) are widely used in the social sciences to measure unobserved constructs of interest using several correlated observed variables. The Generalized Linear Latent Variable Model (GLLVM, Moustaki and Knott, 2000, Skrondal and Rabe-Hesketh, 2004, Bartholomew et al., 2011) is a versatile modeling framework where i) conditional on the latent variables, each observed variable follows a distribution from the exponential family, and ii) only the mean of this conditional distribution depends on the latent variables. The GLLVM encompasses various LVMs for continuous, categorical, and count observed variables, making it a widely used tool for analyzing multivariate data.

While the GLLVM framework is flexible, it can sometimes oversimplify real-world scenarios by relying only on distributions from the exponential family. In specific applications, it is essential to model higher-order moments of the observed variables, such as variance, skewness, and kurtosis, as functions of the latent variables. Ignoring these features of the observed data can result in underestimated standard errors, biased parameter estimates, and inaccurate model fit indices (Lei and Lomax, 2005, Wall et al., 2015, Lai, 2018).

The challenges above have been extensively studied in the literature. Limited information and robust estimation methods have been proposed to address deviations from distributional assumptions (e.g., Browne, 1984, Bollen, 1996, Moustaki and Victoria-Feser, 2006). Furthermore, new advanced models have been developed, such as the heteroscedastic factor models (e.g., Lewin-Koh and Amemiya, 2003, Hessen and Dolan, 2009), LVMs for continuous data displaying skewness and/or kurtosis (e.g., Montanari and Viroli, 2010, Molenaar et al., 2010, Liu and Lin, 2015, Asparouhov and Muthén, 2016), factor models for discrete, count, and bounded continuous data with zero/one/maximum inflation and heaping (e.g., Wang, 2010, Wall et al., 2015, Niku et al., 2017, Magnus and Thissen, 2017, Molenaar et al., 2022), and LVMs for censored/truncated data (e.g., Moustaki and Steele, 2005). However, these models have primarily been developed in isolation and differ in their estimation and inferential methods.

This paper introduces a comprehensive modeling framework called the Generalized Latent Variable Model for Location, Scale, and Shape parameters (GLVM-LSS). This framework models the conditional distribution of each observed variable as a function of latent variables. We achieve this by defining the distributional parameters characterizing each observed variable's conditional distribution as functions of the latent variables. Since the mean and other higher-order moments of the observed variables are expressed in terms of their distributional parameters, they also depend on the latent variables.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

The GLVM-LSS borrows ideas from the Generalized Additive Model for Location, Scale, and Shape (GAMLSS) regression framework (Rigby and Stasinopoulos, 2005, Klein et al., 2015, Umlauf et al., 2018), and applies them to models with latent variables. The GAMLSS is a flexible regression framework where observed covariates have linear or nonlinear effects on the distributional parameters characterizing the distribution of the outcome variable. It can also accommodate spatial, temporal, and random effects. For a comprehensive treatment of the GAMLSS regression framework, see, e.g., Stasinopoulos et al. (2017, 2024), and Rigby et al. (2020). In this paper, we only consider *linear* effects of the latent variables on the distributional parameters.

Our proposed framework shares similarities with previous works in the LVM literature. For example, in multilevel or longitudinal studies (Skrondal and Rabe-Hesketh, 2004, Hedeker and Gibbons, 2006), location-scale mixed-effects models accommodate covariates and random effects on both location and scale parameters for continuous (Hedeker et al., 2008, 2012) and binary/ordered categorical (Greene, 2003, Hedeker et al., 2006, 2009, 2016) observed variables (the latter using an underlying response formulation, see Remark 2). These models can be fitted using commercial software like the `gllamm` module in `Stata` (Rabe-Hesketh et al., 2004) or the `PROC NLMIXED` routine in `SAS`. An important remark is that multiple group LVMs (Davidov et al., 2018) also allow for different (conditional) means and (conditional) variances among the groups.

1.1 Motivating Examples

We present two motivating examples from different fields where modeling higher-order moments of complex multivariate datasets is of substantive interest. These examples are discussed further in Section 3.

Example 1: The first example comes from public opinion research, where we examine people's attitudes towards different social groups using survey data from the 2020 American National Election Study (ANES 2020). Respondents rate their feelings about these groups on a scale from 0 to 100, with higher ratings indicating more favorable opinions. Although the questions are not explicitly designed to measure a particular latent construct, they provide insight into respondents' positions on a *conservative-progressive* belief scale. We could model the conditional mean of these doubly-bounded responses as a function of the latent variable using a Beta factor model (Noel and Dauvier, 2007, Noel, 2014, Revuelta et al., 2022). However, research has shown that liberals are more likely to have similar political attitudes compared to conservatives (Ondish and Stern, 2018), suggesting that the *conservative-progressive* latent factor influences not only the

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

conditional mean but also the conditional variance of the responses. Addressing this issue is possible with a *heteroscedastic* Beta factor model, which we introduce in Section 2.1. Most work on factor models for continuous doubly-bounded observed variables focuses on modeling only the location parameter (conditional mean) in terms of the latent variables. The scale parameter related to the conditional variance does not depend on the latent variables. A notable exception is the mixed and mixture Beta regression model by Verkuilen and Smithson (2012), where the scale parameter of the Beta distribution is modeled using random effects.

Example 2: The second example comes from educational testing, using data from the PISA 2018 computer-based mathematics exam. We present a confirmatory factor model to analyze binary item responses (IRs) and continuous response times (RTs) simultaneously. van der Linden (2007, 2009) proposed a joint model in which the two-parameter logistic model is applied to the IRs, while a linear factor model is used for the log-RTs. To account for the 'speed-accuracy trade-off' in educational testing (Zimmerman, 2011), it is assumed that the latent ability and speed factors are correlated. However, van der Linden's model for the log-RTs (conditional mean) can be restrictive because RTs tend to have a variance that increases with the mean (Van Zandt, 2002, De Boeck and Jeon, 2019). Specifically, the variance parameter, which helps discriminate between test takers with different speed levels, does not depend on the respondent's latent speed trait. Furthermore, modeling additional aspects of the response time distribution as a function of the latent speed trait, such as the (conditional) variance and (conditional) skewness, can provide further insights into individuals' test-taking strategies and items' characteristics. In Section 2.1, we present a joint model for item responses and response times in which log-RTs follow a Skew-Normal distribution (SN, Azzalini, 1985, 2005), with distributional parameters influencing higher order moments modeled as functions of the individual's latent speed trait. Although alternative parametric distributions have been used to model RTs in the hierarchical model framework (e.g., Loeys et al., 2011), the focus is still on the location parameter.

The paper is organized as follows. In Section 2, we introduce the GLVM-LSS model, discuss parameter estimation via full-information marginal maximum likelihood estimation, and examine model identification. In Section 3, we apply the proposed method to real-world data on public opinion research and educational testing. To demonstrate the properties of our proposed method under finite sample settings, we conduct simulation studies in Section 4. Finally, we discuss some limitations and future research directions.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

2. Generalized Latent Variable Models for Location, Scale and Shape parameters (GLVM-LSS)

2.1 Proposed Model Framework

Let $\mathbf{y} = (y_1, ..., y_p)^\intercal$ be a random vector of observed variables with domain $\mathcal{D} \subseteq \mathbb{R}^p$, and $\mathbf{z} = (z_1, ..., z_q)^\intercal \in \mathbb{R}^q$ a random vector of continuous latent variables, with q (much) smaller than p. Assuming local independence (Bartholomew et al., 2011, Chapter 1), the marginal distribution of $\mathbf{y}$ is:

$$f(\mathbf{y}) = \int_{\mathbb{R}^q} \left[\prod_{i=1}^{p} f_i(y_i \,|\, \mathbf{z}; \boldsymbol{\theta}_i) \right] p(\mathbf{z}; \boldsymbol{\Phi}) \, \mathrm{d}\mathbf{z} \,, \tag{1}$$

where, for observed variables $i = 1, \ldots, p$, $f_i(y_i \,|\, \mathbf{z}; \boldsymbol{\theta}_i)$ is the conditional distribution of y_i given $\mathbf{z}$ and $\boldsymbol{\theta}_i = (\theta_i^{(1)}, ..., \theta_i^{(D)})^\intercal$ is a D-dimensional vector of distributional parameters indexing f_i. The (conditional) distributional moments of y_i (mean, variance, skewness) are functions of the parameters $\boldsymbol{\theta}_i$. $p(\mathbf{z})$ is the prior distribution of $\mathbf{z}$, commonly assumed to be a multivariate Normal distribution with covariance matrix $\boldsymbol{\Phi}$, $\mathbf{z} \sim \mathbb{N}(\mathbf{0}, \boldsymbol{\Phi})$.

We propose a class of Generalized Latent Variable Models for Location, Scale and Shape parameters (GLVM-LSS) where the distributional parameters $\theta_i^{(d)} \in \boldsymbol{\theta}_i$ are expressed as monotone functions of linear combinations of the latent variables. We write $\theta_i^{(d)}(\mathbf{z})$ to denote the functional dependence of the distributional parameter on $\mathbf{z}$, but in most cases we omit it to simplify notation. Moreover, we use the sub-index (i, θ_d) to indicate that the corresponding function or model parameter is related to $\theta_i^{(d)}(\mathbf{z})$. The relationship between y_i and $\mathbf{z}$ is therefore through the vector $\boldsymbol{\theta}_i(\mathbf{z})$ in f_i and is determined by the system of equations:

$$v_{i,\theta_d}(\theta_i^{(d)}(\mathbf{z})) = \eta_{i,\theta_d} := \alpha_{i0,\theta_d} + \sum_{j=1}^{q} \alpha_{ij,\theta_d} z_j, \qquad d = 1, ..., D \,, \tag{2}$$

where the parameter-specific link function, denoted by v_{i,θ_d}, is used to ensure that the distributional parameters have appropriate restrictions. The link function can be identity, log, logit, or any other suitable monotone function. η_{i,θ_d} represents the linear combination of latent variables, with intercept α_{i0,θ_d} and factor loadings grouped in the q-dimensional vector $\boldsymbol{\alpha}_{i,\theta_d} = (\alpha_{i1,\theta_d}, ..., \alpha_{iq,\theta_d})^\intercal$.

The distributional parameters $\boldsymbol{\theta}_i$ that describe the shape of f_i can be divided into three categories: location, scale, or shape parameters. Their role depends on which distributional moment of y_i they define. To simplify notation, we refer to the location

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

parameter as $\mu := \theta_i^{(1)}$, the scale parameter as $\sigma_i := \theta_i^{(2)}$, and the shape parameters as $\nu_i := \theta_i^{(3)}$ and $\tau_i := \theta_i^{(4)}$. In most cases, a maximum of four parameters (one location, one scale, and two shape parameters) is enough. However, this framework can be extended to include distributions with multiple location, scale, or shape parameters. We then write $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \nu_i, \tau_i)^\intercal$ to denote the vector of distributional parameters indexing f_i, and use $\varphi_i \in \boldsymbol{\theta}_i$ to refer to any location, scale, or shape parameter in the (conditional) distribution of y_i.

For a more compact matrix notation, let $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_p)^\intercal$ be the vector of the same distributional parameter φ for all y_i's. Denote $\boldsymbol{\alpha}_{0,\varphi} = (\alpha_{10,\varphi}, ..., \alpha_{p0,\varphi})^\intercal$ as a vector of intercepts, and let A_φ be a $(p \times q)$ factor loadings matrix with rows corresponding to the vectors $\boldsymbol{\alpha}_{i,\varphi}$. Finally, let v_φ be the vector function that applies the corresponding link function $v_{i,\varphi}$ to each entry of $\boldsymbol{\varphi}$. Under this convention, the set of equations for a distributional parameter φ is $v_\varphi(\boldsymbol{\varphi}(\mathbf{z})) = \boldsymbol{\alpha}_{0,\varphi} + \mathrm{A}_\varphi \mathbf{z}$.

To further simplify notation, we write the vector of parameters $\boldsymbol{\theta}^\intercal = (\boldsymbol{\mu}^\intercal, \boldsymbol{\sigma}^\intercal, \boldsymbol{\nu}^\intercal, \boldsymbol{\tau}^\intercal)$, the vector of intercepts $\boldsymbol{\alpha}_0^\intercal = (\boldsymbol{\alpha}_{0,\mu}^\intercal, \boldsymbol{\alpha}_{0,\sigma}^\intercal, \boldsymbol{\alpha}_{0,\nu}^\intercal, \boldsymbol{\alpha}_{0,\tau}^\intercal)$, and the factor loading matrix $\mathrm{A}^\intercal = [\mathrm{A}_\mu^\intercal, \mathrm{A}_\sigma^\intercal, \mathrm{A}_\nu^\intercal, \mathrm{A}_\tau^\intercal]$, to compactly express the system of equations of a GLVM-LSS model as:

$$v(\boldsymbol{\theta}(\mathbf{z})) = \boldsymbol{\alpha}_0 + \mathrm{A}\mathbf{z} \tag{3}$$

Remark 1: The GLVM-LSS framework is most useful when observed variables follow distributions with multiple location, scale and shape parameters, and it is appropriate and essential to model their higher-order moments as functions of the latent variables.

Remark 2: While standard LVMs for categorical data fit within the GLVM-LSS framework, accommodating models that involve scale parameters –such as the family of scaled (heteroscedastic) logistic or probit models for binary and ordinal data (see, e.g., Greene, 2003, Hedeker et al., 2006, 2009, 2016, Molenaar et al., 2012, Molenaar, 2015)– requires an alternative model specification. These models employ an *underlying variable* formulation (Jöreskog and Moustaki, 2001) for categorical variables where the observed category denoted by c_i of the ordinal manifest variable y_i is determined by an unobserved *continuous* response $y_i^* \in \mathbb{R}$ underlying the ordinal variable y_i. The connection between y_i and y_i^* is $y_i = c_i \iff \tau_i^{(c_i-1)} < y_i^* \leq \tau_i^{(c_i)}$, where $c_i \in \{1, \ldots, C_i\}$ and $\tau_i^{(0)} = -\infty$, $\tau_i^{(1)} < \cdots < \tau_i^{(C_i-1)}$, $\tau_i^{(C_i)} = +\infty$ are called threshold parameters. Here, $y_i^* \mid \mathbf{z} \sim \mathbb{N}(\mu_i^*(\mathbf{z}), \sigma_i^*(\mathbf{z}))$, and appropriate measurement equations for the location and scale parameters of the underlying response are chosen. Binary variables are special cases. In principle, the underlying response formulation for binary and ordered categori-

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

cal data can be accommodated within the GLVM-LSS framework. However, the current proposed GLVM-LSS framework includes only the homoscedastic logit model. We plan to incorporate the heteroscedastic logit/probit model in future work on the GLVM-LSS framework.

Remark 3: Additional attention is required for discrete observed variables following distributions with distributional parameters taking values in a discrete space (e.g., the natural numbers). In some cases, a reparametrization is available such that the distributional parameters are in the real numbers (see, e.g., Rigby et al. (2020, page 483) for the Negative Binomial distribution case), and common modeling techniques can be used. However, if no alternative parametrization exists, these parameters should be treated as *fixed* and *known*. Otherwise, modeling these distributional parameters requires computational methods that are beyond the scope of this paper (see, e.g., Hammersley, 1950, Choirat and Seri, 2012).

We now revisit the motivating examples in Section 1.1 to illustrate how the GLVM-LSS framework extends the existing LVM literature by modeling features of the observed data beyond the conditional mean:

Example 1 (*continued*). A heteroscedastic Beta factor model: We propose a novel heteroscedastic Beta factor model to model the conditional variance of the doubly-bounded variables in the 2020 ANES dataset. The observed variables conditional on the *conservative-progressive* latent variable (denoted by z) follow a location-scale reparametrization of the Beta distribution, $y_i \mid z \sim \text{Beta}(\mu_i(z), \sigma_i(z))$, where the location parameter $\mu_i(z) \in (0,1)$ and the scale parameter $\sigma_i(z) \in (0,1)$ are modeled as functions of z (Rigby et al., 2020, page 461). Under this parametrization, we have $\mathbb{E}(y_i \mid z) = \mu_i(z)$ and $\text{Var}(y_i \mid z) = \sigma_i^2(z)\mu_i(z)(1-\mu_i(z))$.

This parametrization is closely related to the location-precision parametrization in the heteroscedastic Beta regression literature (Smithson and Verkuilen, 2006, Verkuilen and Smithson, 2012), where a precision parameter $\phi_i(z) \in \mathbb{R}^+$ replaces the scale parameter, and $\text{Var}(y_i \mid z) = (1+\phi_i(z))^{-1}\mu_i(z)(1-\mu_i(z))$. Notably, $\sigma_i^2(z) \equiv (1+\phi_i(z))^{-1}$, meaning $\sigma_i^2(z) \to 0$ when $\phi_i(z) \to \infty$ and $\sigma_i^2(z) \to 1$ when $\phi_i(z) \to 0$. However, the authors of the papers above report computational challenges and biased coefficient estimates in the precision parameter equation. Our empirical application results in Section 3.1 and the simulations study in Section 4.1 show that the location-scale parametrization used in this paper avoids these issues while enabling a direct interpretation of the effect of the latent variables on the conditional variance of the items.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

The equations for the location and scale parameters are:

$$\text{logit}(\mu_i(z)) = \alpha_{i0,\mu} + \alpha_{i1,\mu} z\,, \tag{4}$$

$$\text{logit}(\sigma_i(z)) = \alpha_{i0,\sigma} + \alpha_{i1,\sigma} z\,, \tag{5}$$

with the logit link function mapping the equations onto the correct space for the corresponding distributional parameter.

Example 2 (*continued*). A confirmatory factor model for binary items and skewed response times: The GLVM-LSS specification of the joint model for item responses (IRs) and responses times (RTs) is as follows. For IRs y_i, $i = 1, \ldots, p$, we assume $y_i \mid z_1 \sim \text{Bernoulli}(\pi_i(z_1))$, where z_1 represents the student's latent ability. The equations for the location parameters of the Bernoulli distribution are modeled as:

$$\text{logit}(\pi_i(z_1)) = \alpha_{i0,\pi} + \alpha_{i1,\pi} z_1 \tag{6}$$

where $\alpha_{i0,\pi}$ and $\alpha_{i1,\pi}$ denote item i's difficulty and discrimination parameters, respectively. For the RTs, we assume $\log(t_i) \mid z_2 \sim \text{SN}(\mu_i(z_2), \sigma_i^2(z_2), \nu_i(z_2))$, with t_i denoting item's i RT in minutes and z_2 the student's latent speed trait. Under this reparametrization of the SN distribution[1], the equations for the location $\mu_i(z_2) \in \mathbb{R}$, scale $\sigma_i(z_2) \in \mathbb{R}^+$, and shape $\nu_i(z_2) \in (0, 1)$ parameters are:

$$\mu_i(z_2) = \alpha_{i0,\mu} + \alpha_{i1,\mu} z_2\,, \tag{7}$$

$$\log(\sigma_i(z_2)) = \alpha_{i0,\sigma} + \alpha_{i1,\sigma} z_2\,, \tag{8}$$

$$\text{logit}(\nu_i(z_2)) = \alpha_{i0,\nu} + \alpha_{i1,\nu} z_2\,, \tag{9}$$

respectively, with the identity, log, and logit links mapping the equations onto the respective spaces of the distributional parameters. The (conditional) moments for the log-RTs are $\mathbb{E}(\log(t_i) \mid z_2) = \mu_i(z_2)$, $\text{Var}(\log(t_i) \mid z_2) = \sigma_i^2(z_2)$, and $\text{Skewness}(\log(t_i) \mid z_2) = \tilde{\gamma}(2\,\nu_i(z_2) - 1)$, with $\tilde{\gamma} = \sqrt{2}(4 - \pi) \cdot (\pi - 2)^{-3/2} \approx 0.9953$. Finally, to capture the 'speed-accuracy trade-off', the latent ability and speed traits are distributed as $(z_1, z_2)^\intercal \sim \mathbb{N}_2(\mathbf{0}, \mathbf{\Phi})$, where $\mathbf{\Phi}$ is a correlation matrix.

[1]For modeling convenience, the shape parameter is restricted to the $(0, 1)$ interval. We achieve this by applying a monotonic transformation of the shape parameter in the 'centered' parametrization outlined in Azzalini (1985) and described in Azzalini and Capitanio (1999). We refer the readers to Online Supplementary Materials A1 for further details.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Remark 4: The proposed framework can be extended to accommodate linear and non-linear structural relationships between latent variables and/or observed covariates effects. Let $\mathbf{x} = (x_1, \ldots, x_s)^\intercal \in \mathbb{R}^s$ denote a vector of observed covariates. A linear structural model with covariate effects can be written in matrix form as:

$$\mathbf{z} = \mathbf{G}\mathbf{z} + \mathbf{B}\mathbf{x} + \boldsymbol{\delta}\,,$$

where $\mathbf{G}$ is a $q \times q$ matrix of structural parameters satisfying recursive restrictions (Bollen, 1989), $\mathbf{B}$ is a $q \times s$ matrix of regression coefficients, and $\boldsymbol{\delta} \sim \mathbb{N}(\mathbf{0}, \mathbb{I}_q)$ is a vector of independent standard Normal errors.

Similarly, assuming the vector of covariates $\mathbf{x}$ has a linear and additive effect on the linear component, the measurement equation for an arbitrary distributional parameter $\varphi_i \in \boldsymbol{\theta}_i$ indexing $f_i(y_i \,|\, \mathbf{z}, \mathbf{x}; \boldsymbol{\theta}_i)$ becomes:

$$v_{i,\varphi}(\varphi_i(\mathbf{z}, \mathbf{x})) = \alpha_{i0,\varphi} + \sum_{j=1}^{q} \alpha_{ij,\varphi} z_j + \sum_{r=1}^{s} \beta_{ir,\varphi} x_r\,, \tag{10}$$

where $\boldsymbol{\beta}_{i,\varphi} = (\beta_{i1,\varphi}, \ldots, \beta_{is,\varphi})^\intercal$ is a vector of regression coefficients. The measurement equation in (10) can also include interactions between the observed covariates and the latent variables to enable the study of 'distributional' differential item functioning (DIF), where the assessment of DIF goes beyond the (conditional) mean and extends to the items' (conditional) higher-order moments.

While Remark 4 highlights how the GLVM-LSS can be expanded into a more general LVM framework, this paper aims to establish a foundation for future methodological developments in distributional LVM research. It should be noted, however, that the current implementation of the GLVM-LSS improves model fit over traditional approaches and proves valuable in empirical research where, from a measurement perspective, higher-order moments of observed variables carry substantive meaning or reflect important item characteristics.

For example, in the ANES 2020 application (Section 3.1), we examine whether individuals with liberal values have more homogeneous views on the social groups in question. If so, the (conditional) variance of the thermometer items should be lower for individuals on the *liberal* side of the latent scale than for those on the *conservative* side. In the PISA 2018 application (Section 3.2), modeling the scale (variance) and shape (skewness) parameter of the (log-)RTs as functions of the latent speed factor could help testing agencies better understand test-taking strategies and item characteristics.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

2.2 Model Identification

The generality of the proposed model makes it challenging to derive general conditions for global identification (e.g., Skrondal and Rabe-Hesketh, 2004, Chapter 5), so we instead resort to the weaker notion of (strict) local identification (Rothenberg, 1971).

As in the GLLVM, strict local identifiability in the GLVM-LSS is only possible for points on the *reduced* parameter space that results from imposing at least q^2 restrictions across the factor loadings matrix A and the latent variables covariance matrix $\mathbf{\Phi}$. These restrictions address rotational indeterminacy and fix the scale of the latent variable space (Anderson and Rubin, 1956). Denote by Ξ such *reduced* parameter space and $\Theta \in \Xi$ a point of free (unrestricted) model parameters. A point $\Theta_0 \in \Xi$ is strictly locally identifiable if the expected information matrix $\mathcal{I}(\Theta_0)$ is strictly positive definite (Rothenberg, 1971). However, in the GLVM-LSS framework, model identifiability can be challenging even after imposing appropriate restrictions on the parameters due to the presence of multiple (possibly correlated) location, scale, and shape parameters. In what follows, we present theoretical results on the identifiability of GLVM-LSS models for continuous observed data following distributions with multiple distributional parameters.

Under suitable regularity conditions, we show that GLVM-LSS is generically locally identifiable in the sense that the model is strictly locally identifiable for almost all parameters in Ξ except for a subset with Lebesgue measure zero. More precisely, we define 'generic local identifiability' as follows:

Definition 2.1. *A statistical model is generically locally identified if every $\Theta \in \Xi \setminus V$ is (strictly) locally identifiable, where V is a proper sub-variety of Ξ and thus has Lebesgue measure zero in Ξ.*

This notion of identifiability is closely related to and can be seen as a weaker version of the concept of 'generic identifiability' (Allman et al., 2009, Gu and Xu, 2020) that has been commonly adopted for studying the identifiability of latent variable models. The following Theorem holds:

Theorem 2.1. *Assume:*

(A1) There exists a point in the reduced parameter space $\Theta_0 \in \Xi$ such that $\mathcal{I}(\Theta_0)$ is strictly positive definite,

(A2) $f_i(y_i \mid \mathbf{z}; \boldsymbol{\theta}_i(\Theta))$, $i = 1, \ldots, p$, in the measurement part, and $p(\mathbf{z}; \Theta)$ in the structural part of a GLVM-LSS model, are infinitely differentiable in Ξ and $\mathcal{D}$, and their respective supports are independent of Θ.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Then, the GLVM-LSS model is generically locally identified.

Assumption (A1) avoids trivial non-identification issues, and follows from the restrictions imposed to solve the rotational and scale indeterminacies. This assumption also rules out cases where the distributional parameters in $\boldsymbol{\theta}_i$ are linearly dependent. Assumption (A2) relates to smoothness and regularity conditions often met in practice for continuous distributions indexed by multiple location, scale, and shape parameters. The proof of Theorem 2.1, and auxiliary definitions, lemmas, and propositions are provided in the Online Supplementary Material A3.

While Theorem 2.1 addresses models with continuous observed data, establishing general identification conditions for GLVM-LSS models with categorical data is more challenging due to the finite amount of information in the data. In these cases, parameter identification follows from the existence of a finite-dimensional sufficient statistic. Indeed, once appropriate parameter restrictions are imposed, a necessary condition for a GLVM-LSS with categorical items to be identified is that the number of parameters is less than the number of possible response patterns (e.g., 2^p for binary items or $\prod_{i=1}^{p} C_i$ for categorical items, where C_i is the number of categories for item i). We note that, as mentioned in Remark 2, the current specification of the proposed framework does not cover heteroscedastic models for binary/ordinal categorical data and therefore the identification constraints described in, e.g., Skrondal and Rabe-Hesketh (2004, Chapter 2), Hedeker et al. (2006, 2009) or Molenaar et al. (2012), Molenaar (2015), do not apply to the current setting.

In practice, some f_i's might be indexed by distributional parameters that are correlated (yet linearly independent). Due to sampling variability, the latter can lead to situations where the model is not empirically identified. In this case, empirical local identification of the MLE can be verified if the estimated expected information matrix, $\hat{\mathcal{I}}(\hat{\Theta})$, is non-singular (McDonald and Krane, 1977).

2.3 Parameter Estimation and Computation

For a random sample of n independent observations, the marginal log-likelihood is:

$$\ell(\Theta; \mathbf{Y}) = \sum_{m=1}^{n} \log \left(\int_{\mathbb{R}^q} \left[\prod_{i=1}^{p} f_i(y_{mi} \,|\, \mathbf{z}; \boldsymbol{\theta}_i(\boldsymbol{\alpha}_0, \mathrm{A})) \right] p(\mathbf{z}; \boldsymbol{\Phi}) \, \mathrm{d}\mathbf{z} \right) , \tag{10}$$

where $\mathbf{Y} \in \mathbb{R}^{n \times p}$ is the observed data matrix with rows given by the p-dimensional vectors of observed variables $\mathbf{y}_m = (y_{m1}, \ldots, y_{mp})^\intercal$ from units $m = 1, \ldots, n$, and Θ is a

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

K-dimensional vector of unknown model parameters. For computational convenience, we parameterize the factor covariance matrix through its Cholesky decomposition $\boldsymbol{\Phi} = \mathrm{L}\mathrm{L}^\intercal$, where L is a lower triangular matrix. When the latent variables are uncorrelated, L is fixed to the identity matrix, and $\Theta^\intercal = (\boldsymbol{\alpha}_0^\intercal, \mathrm{vec}(\mathrm{A})^\intercal)$, where 'vec' is the vectorization operator that concatenates the free entries of A in a vector. In confirmatory settings, where the latent variables are correlated, $\Theta^\intercal = (\boldsymbol{\alpha}_0^\intercal, \mathrm{vec}(\mathrm{A})^\intercal, \mathrm{vech}(\mathrm{L})^\intercal)$, where 'vech' is the half-vectorization operator that concatenates the free lower-triangular entries of L in a vector.

The model parameters are estimated via full-information marginal maximum likelihood (MML). Let $\Xi \subseteq \mathbb{R}^K$ be the reduced parameter space. The maximum likelihood estimate (MLE), denoted by $\hat{\Theta}$, is:

$$\hat{\Theta} = \underset{\Theta \in \Xi}{\arg\max}\, \ell(\Theta; \mathbf{Y})$$

To compute the MLE, we find the solution to the system of (non-linear) score equations $\mathbb{S}(\Theta; \mathbf{Y}) := \nabla_\Theta \ell = \mathbf{0}$, with entries of the general form

$$\mathbb{S}_{i,\varphi_i}(\Theta; \mathbf{Y}) := \frac{\partial \ell(\Theta; \mathbf{Y})}{\partial \boldsymbol{\alpha}_{i,\varphi}} = \sum_{m=1}^{n} \int_{\mathbb{R}^q} \left[\frac{\partial \log f_i(y_{im} \,|\, \mathbf{z})}{\partial \varphi_i} \frac{\partial \varphi_i}{\partial \eta_{i,\varphi}} \frac{\partial \eta_{i,\varphi}}{\partial \boldsymbol{\alpha}_{i,\varphi}} \right] p(\mathbf{z} \,|\, \mathbf{y}_m; \Theta)\, \mathrm{d}\mathbf{z} \quad (11)$$

for the vector of factor loadings $\boldsymbol{\alpha}_{i,\varphi}$ in the measurement equation of the distributional parameter $\varphi_i \in \boldsymbol{\theta}_i$, $i = 1, \ldots, p$. We provide expressions of the score equations (11) for the GLVM-LSS models introduced in this paper in the Online Supplementary Material A1. Scores for the $[j,k]^{\text{th}}$ entry in L are:

$$\mathbb{S}_{\mathrm{L}_{[j,k]}}(\Theta; \mathbf{Y}) := \frac{\partial \ell(\Theta; \mathbf{Y})}{\partial \mathrm{L}_{[j,k]}} = -n \,\mathrm{tr}\left(\mathrm{L}^\intercal (\mathrm{L}\mathrm{L}^\intercal)^{-1} \mathrm{D}_{jk}\right) + \sum_{m=1}^{n} \left[\mathrm{tr}\left(\mathrm{G}_{jk} \mathbb{V}_m\right) + \breve{\mathbf{z}}_m^\intercal \mathrm{G}_{jk} \breve{\mathbf{z}}_m\right] \quad (12)$$

where $\mathrm{D}_{jk} = \partial \mathrm{L} / \partial \mathrm{L}_{[j,k]}$ is a square matrix of dimension q, with a value of 1 in the $[j,k]$ position and zero elsewhere; $\mathrm{G}_{jk} = (\mathrm{L}\mathrm{L}^\intercal)^{-1} \mathrm{D}_{jk} \mathrm{L}^\intercal (\mathrm{L}\mathrm{L}^\intercal)^{-1}$; and the conditional mean $\breve{\mathbf{z}}_m = \mathbb{E}(\mathbf{z} \mid \mathbf{y}_m; \Theta)$ and conditional variance $\mathbb{V}_m = \mathbb{E}((\mathbf{z} - \breve{\mathbf{z}}_m)(\mathbf{z} - \breve{\mathbf{z}}_m)^\intercal \mid \mathbf{y}_m; \Theta)$ are obtained using the properties of the trace operator and the linearity of the conditional expectation. Details on the computation of L are discussed in the Online Supplementary Material A2.

In most cases, $\hat{\Theta}$ is computed using iterative score-based optimization algorithms. Our estimation strategy begins with a warm-up phase using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977, Bock and Aitkin, 1981), followed by a direct maximization of the marginal log-likelihood with the BFGS quasi-Newton algorithm (No-

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

cedal and Wright, 2006, Chapter 6). The transition between these two algorithms is possible due to the equivalence of the score functions of the complete-data and the marginal log-likelihoods (Louis, 1982).

Upon computation of the MLE, in exploratory settings, an orthogonal or oblique rotation can be applied to the estimated factor loading matrix, $\hat{\mathbf{A}}$, to obtain a more interpretable and sparse solution (e.g., Jennrich, 2004, 2006, Liu et al., 2023).

Unless the objective function is strictly concave, both the (quasi-)Newton update in the EM algorithm's M-step and the BFGS algorithm converge to a *local* maximum. This means that the final solution can depend on the initial values chosen. Using different starting values and comparing the resulting marginal log-likelihood estimates is advisable to determine the best solution.

For the one- and two-dimensional models presented in Sections 3 and 4, it is sufficient to evaluate the integrals in the score vector and the information matrices numerically using an ordinary Gauss-Hermite (GH) rule. This can be done with a fixed number of user-defined quadrature points on each dimension of the space of the latent variables. However, the GH approach runs into computational challenges when the dimension of the latent space is high. In such cases, alternatives like adaptive GH quadrature (Rabe-Hesketh et al., 2005, Schilling and Bock, 2005) or stochastic approximation methods (Cai, 2010, Zhang and Chen, 2022) should be considered.

For model selection, we suggest using information criteria for nested (e.g., with restrictions on the model parameters) and non-nested (e.g., GLVM-LSS models with different distributions on the measurement model) models. The Akaike Information Criterion (AIC, Akaike, 1974) and the Bayesian Information Criterion (BIC, Schwarz, 1978) are popular criteria to evaluate model fit. The AIC and BIC often concur and are commonly used together in applied research (Kuha, 2004). However, in case of divergent results, we suggest referring to the AIC when dimensionality reduction is the primary goal, as it favors (expected) predictive performance (Shao, 1997); while using the BIC when the study involves substantive interpretation of the latent variables, as it favors consistent model selection (Nishii, 1984).

3. Empirical Applications

We present two empirical applications that follow from the GLVM-LSS examples introduced in Section 2.1.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

3.1 ANES 2020: 'Thermometer' variables

The ANES 2020 dataset contains 'feeling thermometer' variables on social groups, including sexual orientation and gender identity groups (Gay men and Lesbians, Transgender people), social and political movements (Feminists, #MeToo and BLM movements), and groups that were trending in the news during 2020 (labor unions, journalists, scientists) from the post-election sample of the 2020 American National Election Study (ANES)[2]. Participants rate their feelings towards social groups on a scale of 0 to 100, where higher ratings indicate more favorable attitudes.

The ANES thermometer variables have been used in some studies as a substitute for political orientation and measures of personal and societal values (e.g., Abelson et al., 1982, Krasa and Polborn, 2014, Guth, 2019). Similarly, they provide insights into an individual's position on a *conservative-progressive* belief scale. We present results for the one-factor Beta model[3].

We model the observed variables using the heteroscedastic Beta factor model discussed in Section 2.1, i.e., $y_i \,|\, z \sim \text{Beta}(\mu_i(z), \sigma_i(z))$, $i = 1, \ldots, 8$. We scale the responses by $1/100$ so the observed variables are within the interval $(0, 1)$. We also replace extreme responses on the boundaries of the interval with numerical values that are arbitrarily close to 0 and 1 (1^{-9} and $(1 - 1^{-9})$, respectively). We exclude individuals with incomplete interviews or technical errors in their answers from the analysis, treat responses of *'Don't know'*, *'Don't recognize'*, and *'Refuse'* as missing data. The resulting sample consists of 7253 respondents.

The Online Supplementary Material A4a presents descriptive statistics for the observed variables. Most variables have negatively skewed marginal empirical distributions and negative excess kurtosis, except item *Scientists*. The empirical cumulative distribution functions (ECDF) for the observed variables are displayed in Figure 1. Although the thermometer ratings are measured on a continuous scale, respondents tend to round their answers to the nearest 5 or 10, resulting in a stepped appearance in the ECDFs. Some questions show a higher frequency of extreme responses (either zero or one). Most responses tend to cluster around 0.5, suggesting that respondents are reluctant to take a clear position on issues related to the conservative-progressive belief spectrum. We esti-

[2]American National Election Studies, 2021. (www.electionstudies.org). Full Release (dataset and documentation). July 19, 2021 version. These materials are based on work supported by the National Science Foundation under grant numbers SES-1444721, 2014-2017, the University of Michigan, and Stanford University.

[3]We also explored two-factor Beta models, but careful analysis of the expected information matrix (evaluated at the MLE) reveals that the heteroscedastic Beta factor model with $q = 2$ is not of full rank. Thus, the model is not empirically identified for this dataset.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

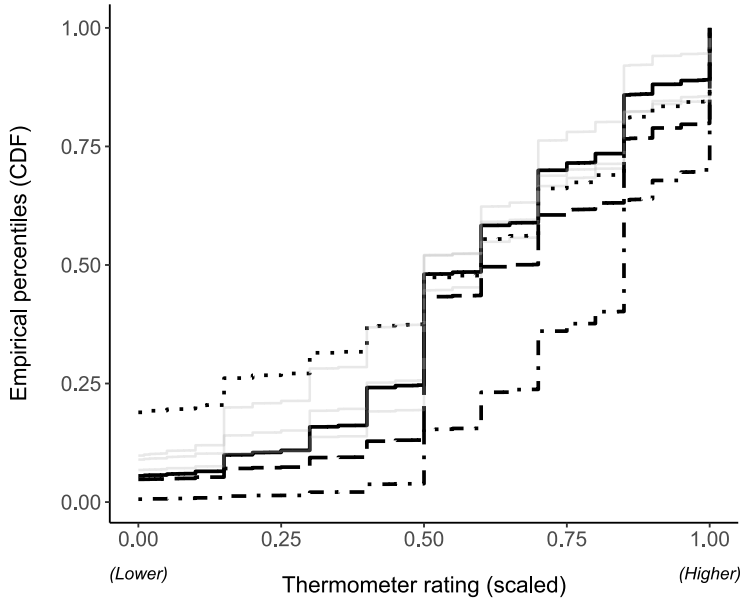


Figure 1: ANES 2020: Empirical cumulative distribution function (ECDF). Highlighted variables: *Feminists* (solid line, ——), *Gay men and Lesbians* (dashed line, - - -), *BLM movement* (dotted line, ·····), and *Scientists* (dash-dot line, -·-·).

mated a baseline (homoscedastic) model, which assumes a constant scale parameter, and an alternative (heteroscedastic) model, which allows the scale parameter to vary based on the latent variable. The heteroscedastic model was selected using the AIC and BIC criteria, as detailed in Table 1. Additionally, Table 2 presents the parameter estimates along with their corresponding estimated standard errors for the heteroscedastic model.

The initial values for the estimation algorithm were chosen by conducting a principal component analysis on the observed data matrix. We then used the first principal component as the explanatory variable in a series of independent distributional regression analyses with a Beta distribution for the outcome variable. To explore potential local solutions, we tested various random starting values; however, the results remained consistent with those reported below.

Model	AIC	BIC	K
Beta $(\mu(z), \sigma)$	-270078.20	-269912.86	24
Beta $(\mu(z), \sigma(z))$	**-271002.07**	**-270781.62**	32

Table 1: ANES 2020: AIC and BIC for the homoscedastic and heteroscedastic Beta factor models. $K = \dim(\hat{\Theta})$ is the number of parameters in the corresponding model.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Item	Location parameter (μ) measurement equation				Scale parameter (σ) measurement equation			
	$\hat{\alpha}_{i0,\mu}$		$\hat{\alpha}_{i1,\mu}$		$\hat{\alpha}_{i0,\sigma}$		$\hat{\alpha}_{i1,\sigma}$	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Gay men and Lesbians	1.00	(0.03)	1.54	(0.02)	0.52	(0.01)	-0.25	(0.01)
Transgender people	0.61	(0.03)	1.70	(0.02)	0.36	(0.01)	-0.15	(0.01)
Feminists	0.49	(0.02)	1.56	(0.02)	0.29	(0.01)	-0.14	(0.01)
#MeToo movement	0.47	(0.03)	1.65	(0.03)	0.57	(0.01)	-0.28	(0.01)
BLM movement	0.12	(0.02)	1.32	(0.03)	1.28	(0.01)	-0.25	(0.01)
Labor Unions	0.46	(0.02)	0.83	(0.01)	0.77	(0.01)	-0.16	(0.01)
Journalists	-0.06	(0.02)	1.21	(0.02)	0.71	(0.01)	-0.16	(0.01)
Scientists	1.72	(0.02)	0.84	(0.02)	0.71	(0.01)	-0.19	(0.01)

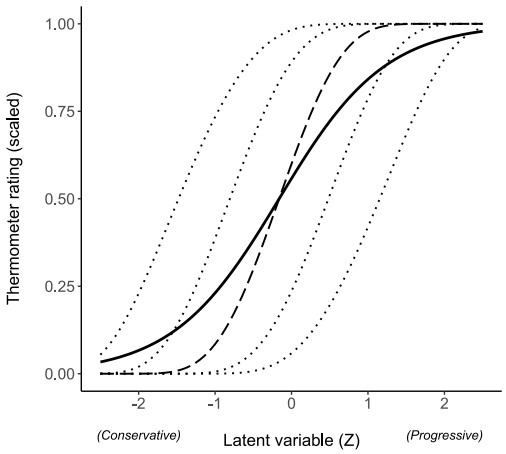
Table 2: ANES 2020: Estimated (Est.) coefficients and their standard errors (SE) for the heteroscedastic Beta factor model.

We first discuss the results of the location parameter equations. The estimated slopes ($\hat{\alpha}_{i1,\mu}$'s) are positive and statistically significant, indicating that more progressive individuals tend to rate these groups higher on average. In addition, lower intercept values ($\hat{\alpha}_{i0,\mu}$'s) indicate that the items are perceived as more challenging, meaning a more progressive position on the latent scale is necessary to achieve at least 50% on these items.

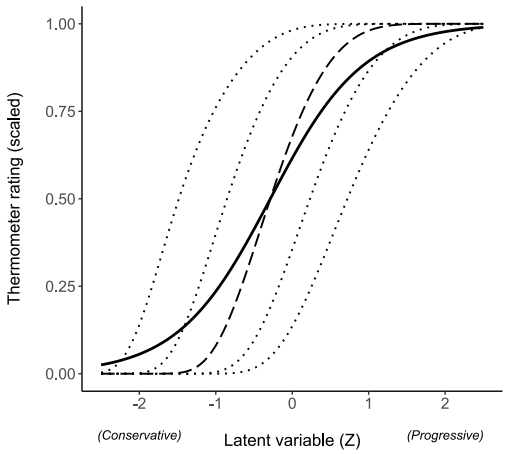
In discussing the scale parameter equations, all the estimated slopes ($\hat{\alpha}_{i1,\sigma}$'s) are negative and statistically significant. This indicates that individuals on the 'progressive' side of the latent scale tend to hold more homogeneous views about these groups than those on the 'conservative' side, in line with previous findings in public opinion research literature. The estimated intercepts ($\hat{\alpha}_{i0,\sigma}$'s) determine the conditional variance for the 'average' position on the latent scale (i.e., $z = 0$). Larger intercepts imply greater heterogeneity in people's responses near the middle point of the latent scale.

A comparison of the fitted (conditional) distribution implied by the homoscedastic and heteroscedastic models for selected variables is shown in Figure 2. The plot includes the fitted mean, median, and percentiles (10th, 25th, 75th, and 90th) for both models. The homoscedastic model (shown in Figures 2a and 2c) does not effectively capture the asymmetries in the conditional distributions of the observed variables along the latent scale. In contrast, the heteroscedastic model, illustrated in Figures 2b and 2d, successfully captures these asymmetries.

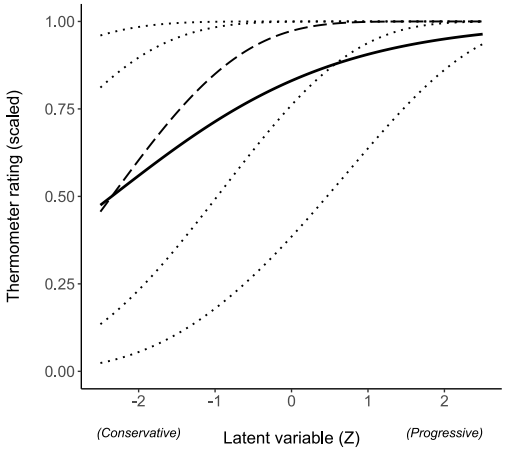
https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press



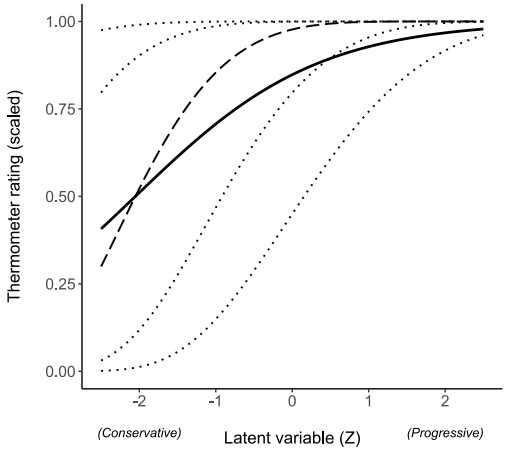
(a) Item: #MeToo movement (homos. model)



(b) Item: #MeToo movement (heter. model)



(c) Item: Scientists (homoscedastic model)



(d) Item: Scientists (heteroscedastic model)

Figure 2: ANES 2020: Fitted conditional expected values (solid line, ——), median (dashed line, - - -), and percentiles (dotted lines, ·····).

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

3.2 PISA 2018: A joint model for item responses and response times

The second dataset was obtained from the 2018 PISA computer-based mathematics exam. We focused on a sample of Brazilian students who answered nine binary items from the first testing booklet. Only individuals who provided complete responses were included, resulting in a final sample size of 1,280 students. Response times for each binary item were recorded in logarithmic minutes. Descriptive statistics for the item responses and log response times are available in the Online Supplementary Material A4b.

Response times (RTs) provide valuable information about a student's ability and test-taking strategies and are also helpful in item calibration and test design (van der Linden, 2007, 2008, van der Linden and Guo, 2008, van der Linden et al., 2010). A hierarchical model for speed and accuracy on test items was initially proposed in van der Linden (2007, 2009), and later extended in Molenaar et al. (2015), Bolsinova et al. (2017), Bolsinova and Molenaar (2018) and others. For a review of models involving items and response times, see De Boeck and Jeon (2019).

The baseline model in van der Linden (2007) assumes that the log-RTs follow a Normal distribution, with only the conditional mean (location parameter) depending on the latent speed factor. Factor loadings for log-RTs are fixed[4], but the variance of the latent speed factor is freely estimated. We extend this model by employing the Skew-Normal (SN) distribution, which models varying heterogeneity and skewness in the log-RTs along the latent speed factor as described in Section 2.1. Variances for the latent ability and latent speed factors are fixed at 1, while the factor loadings are freely estimated. Higher-order moments of RTs can provide valuable insight into students' test-taking strategies, thought processing during high-stakes standardized tests, and information on item quality.

The empirical and model-implied marginal distributions of the response times in log minutes are displayed in Figure 3. We observed that the majority of the log-RTs showed some degree of skewness. The model fit improved when the log-RTs were assumed to follow an SN distribution (solid line) rather than a Normal distribution (dashed line).

We estimated seven increasingly complex models, including the proposed confirmatory factor model discussed in Section 2.1, using the full-information maximum likelihood procedure described in Section 2.3. We tried different starting values to check for local solutions, and the results remained consistent across estimations. We began the estimation process by implementing a warm-start strategy, which involved a PCA decomposition on the matrix of observed variables. We then retained $q = 2$ principal components to use them as observed covariates in a series of distributional regressions with IRs and log-

[4]Slopes in the log-RTs equations are implicitly fixed to -1 in van der Linden's hierarchical model.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

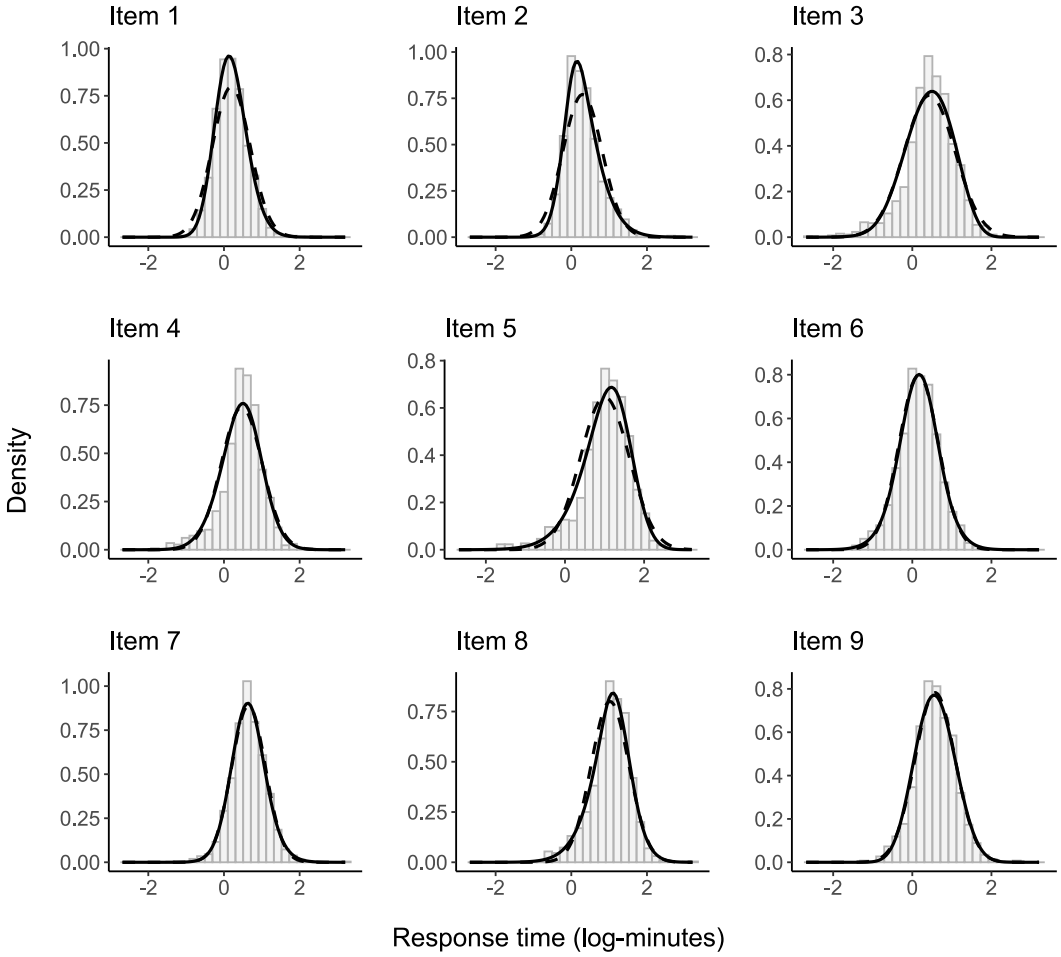


Figure 3: PISA 2018: Empirical and model-implied marginal distributions for response times (in log-minutes). The solid line (——) is the SN model, and the dashed line (- - -) the Normal model.

RTs as outcomes. The two-dimensional integrals were numerically evaluated using the GH quadrature, with 45 quadrature points in each latent dimension (a total of 2025 quadrature points).

Models 1 to 3 assume the log-RTs follow a conditional Normal distribution. Model 1 serves as the baseline hierarchical model described in van der Linden (2007). Model 2 allows to freely estimate the factor loadings in the log-RT model, similar to the 'unrestricted model' in Molenaar et al. (2015), while fixing the variance of the latent speed factor to $\text{Var}(z_2) = 1$ for identification purposes. Model 3 is a heteroscedastic version of Model 2. In models 4 to 7, we assume that the log-RTs follow a conditional Skew-Normal distribution. In Model 4, only the location parameter (μ) depends on the latent speed factor, while the scale (σ) and shape (ν) parameters remain constant. Models 5 and 6 model $(\mu, \sigma)^\intercal$ and $(\mu, \nu)^\intercal$ as functions of z_2, respectively. Model 7, the full-Skew-Normal model, treats all distributional parameters as functions of the latent speed trait. In all

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

cases, the IRT model for the item responses (IRs) remains consistent with what was described above, with the equation given in (6). Results are presented in Table 3. Notably, models with Skew-Normal log-RTs demonstrate better model fit compared to those with Normal log-RTs. Model 7 provides the best fit based on its AIC and BIC values.

Model	AIC	BIC	K
1. Bernoulli ($\pi(z_1)$) + Normal ($\mu(z_2)$, fixed $\alpha_{i1,\mu}$'s)	26173.61	26369.49	38
2. Bernoulli ($\pi(z_1)$) + Normal ($\mu(z_2)$)	25908.68	26145.79	46
3. Bernoulli ($\pi(z_1)$) + Normal ($\mu(z_2), \sigma(z_2)$)	25754.91	26038.42	55
4. Bernoulli ($\pi(z_1)$) + Skew-Normal ($\mu(z_2)$)	25326.08	25609.58	55
5. Bernoulli ($\pi(z_1)$) + Skew-Normal ($\mu(z_2), \sigma(z_2)$)	25281.44	25611.33	64
6. Bernoulli ($\pi(z_1)$) + Skew-Normal ($\mu(z_2), \nu(z_2)$)	25232.76	25562.66	64
7. Bernoulli ($\pi(z_1)$) + Skew-Normal ($\mu(z_2), \sigma(z_2), \nu(z_2)$)	**25172.12**	**25548.41**	73

Table 3: PISA 2018: AIC and BIC for GLVM-LSS for the joint modeling of item responses and response times. $K = \dim(\hat{\Theta})$ is the number of parameters in the corresponding model.

Estimates for the intercepts, loadings, and factor correlation in Model 7, along with their corresponding estimated standard errors, are presented in Table 4. The interpretation of the intercepts and slopes in the equations for the location parameter of the IRs (π_i's) and log-RTs (μ_i's) is straightforward. The $\hat{\alpha}_{i0,\pi}$'s and $\hat{\alpha}_{i1,\pi}$'s represent the difficulty and discrimination parameters for the IRs. In other words, items with lower $\hat{\alpha}_{i0,\pi}$ values are considered more difficult, while those with higher $\hat{\alpha}_{i1,\pi}$ values are seen as having more discrimination power.

The $\hat{\alpha}_{i0,\mu}$'s in log-RTs represent the average log-RTs for $z_2 = 0$, also known as the item's average time intensity (van der Linden, 2007). The estimated slopes $\hat{\alpha}_{i1,\mu}$ in the equation for the location parameter of the log-RTs are all negative. This suggests that individuals with a higher latent speed trait will respond faster to any given item.

The estimated correlation between the latent ability and the speed factor is -0.28 (SE 0.04), suggesting that test takers with higher latent ability generally take longer to respond. This result aligns with previous studies on the speed-accuracy trade-off, which indicates that individuals who respond slowly make fewer mistakes compared to those who respond quickly and make more mistakes (e.g., van der Linden (2007), and Heitz (2014) for a general overview on the subject). Previous studies have found correlations between the latent ability and the latent speed trait of similar sign and magnitude in large-scale educational testing of quantitative subjects (e.g., van der Linden and Guo, 2008).

The equations for the scale (standard deviation) and shape (skewness) parameters

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Item	Estimated coefficients in the equation for item responses (IR)					Estimated coefficients in the equations for (log-)response times (log-RT)											
	Location parameter (π_i)					Location parameter (μ_i)				Scale parameter (σ_i)				Shape parameter (ν_i)			
	$\hat{\alpha}_{i0,\pi}$		$\hat{\alpha}_{i1,\pi}$			$\hat{\alpha}_{i0,\mu}$		$\hat{\alpha}_{i1,\mu}$		$\hat{\alpha}_{i0,\sigma}$		$\hat{\alpha}_{i1,\sigma}$		$\hat{\alpha}_{i0,\nu}$		$\hat{\alpha}_{i1,\nu}$	
	Est.	SE	Est.	SE		Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Item 1	0.64	(0.07)	0.79	(0.10)	0.19	(0.01)	-0.17	(0.01)	-0.95	(0.02)	-0.03	(0.02)	0.63	(0.14)	-0.03	(0.16)	
Item 2	-0.47	(0.07)	1.03	(0.11)	0.30	(0.01)	-0.24	(0.01)	-0.89	(0.02)	-0.10	(0.02)	1.56	(0.16)	-0.78	(0.19)	
Item 3	-0.04	(0.10)	1.94	(0.21)	0.42	(0.02)	-0.25	(0.02)	-0.61	(0.02)	-0.08	(0.02)	-1.07	(0.14)	0.81	(0.16)	
Item 4	-0.69	(0.08)	0.96	(0.11)	0.45	(0.02)	-0.34	(0.01)	-0.87	(0.02)	-0.05	(0.02)	-1.45	(0.14)	-0.33	(0.16)	
Item 5	-2.85	(0.23)	2.28	(0.25)	1.00	(0.02)	-0.36	(0.02)	-0.68	(0.02)	0.14	(0.02)	-1.04	(0.14)	-0.32	(0.21)	
Item 6	-0.91	(0.07)	0.32	(0.08)	0.16	(0.02)	-0.36	(0.01)	-0.97	(0.02)	0.03	(0.03)	0.11	(0.13)	-0.88	(0.22)	
Item 7	-4.84	(0.45)	2.52	(0.34)	0.65	(0.01)	-0.33	(0.01)	-1.15	(0.02)	-0.04	(0.02)	0.35	(0.15)	-0.58	(0.18)	
Item 8	-3.64	(0.32)	2.36	(0.29)	1.02	(0.02)	-0.39	(0.01)	-1.02	(0.02)	0.12	(0.03)	-1.22	(0.22)	-1.61	(0.30)	
Item 9	-2.73	(0.17)	1.45	(0.16)	0.58	(0.02)	-0.30	(0.01)	-0.90	(0.02)	-0.00	(0.02)	0.30	(0.10)	0.36	(0.14)	
Estimated latent correlation (z_1, z_2): $\hat{\phi}_{\mathbf{z}} = -0.28$ (SE: 0.04)																	

Table 4: PISA 2018: Estimated coefficients (Est.) and Standard Errors (SE) for a joint model of item responses and response times (Model 7).

of the log-RTs provide valuable information about the items and RTs. The estimates $\hat{\alpha}_{i1,\sigma}$ and $\hat{\alpha}_{i1,\nu}$ indicate that some log-RTs exhibit heteroscedasticity (items 2, 3, 4, 5, 8) and varying skewness (items 2, 3, 4, 6, 7, 8, 9) in their log-RTs as the latent speed factor changes. Selected item characteristic curves (ICC) for the IRs and the fitted Skew-Normal conditional distributions for the log-RTs (parameterized by the coefficients in Table 4) are shown in Figures 4 and 5. The (conditional) mean, median, and percentiles (0.025, 0.10, 0.25, 0.75, 0.90, and 0.975) for the log-RTs are plotted to illustrate how the distribution's shape changes as the latent speed factor dimension varies.

Figures 4a and 4b illustrate item 2, while Figures 4c and 4d depict item 3. The figures show how the variance of $\log(t_2)$ and $\log(t_3)$ changes as we move along the latent speed factor dimension—the conditional skewness, however, changes in opposite directions. For instance, $\log(t_2)$ is positively skewed for individuals in the left tail of the latent speed factor, while $\log(t_3)$ is negatively skewed for the same group of students. On the other hand, the response times' distributions are symmetric for individuals on the right tail of the speed factor dimension. Figures 5a and 5b depict item 5, while Figures 5c and 5d correspond to item 8. The estimated positive slope for the scale parameter suggests a larger variance in the response times for individuals on the upper tail of the latent speed factor distribution. These items are among the most difficult ones, with higher $\hat{\alpha}_{i0,\pi}$ values, and also require more time on average, with higher $\hat{\alpha}_{i0,\mu}$ values. Furthermore, they also exhibit varying skewness parameters. For example, for item 8, the direction of the skewness changes depending on the location along the latent speed factor scale. These results might suggest differences in item characteristics, such as the wording, task, difficulty, or cognitive processes required for their completion.

4. Simulation Studies

We performed several simulation studies to evaluate the accuracy of the parameter estimates obtained through the MML algorithm explained in Section 2.3, along with their corresponding standard errors. Simulations were conducted in `R` (R Core Team, 2022) using the package `glvmlss`, with underlying functions programmed in `C++` using packages `Rcpp` (Eddelbuettel et al., 2023a), `RcppArmadillo`, (Eddelbuettel et al., 2023b), and `RcppEnsmallen` (Balamuta and Eddelbuettel, 2018). Code and replication files are available at https://github.com/ccardehu/glvmlss.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

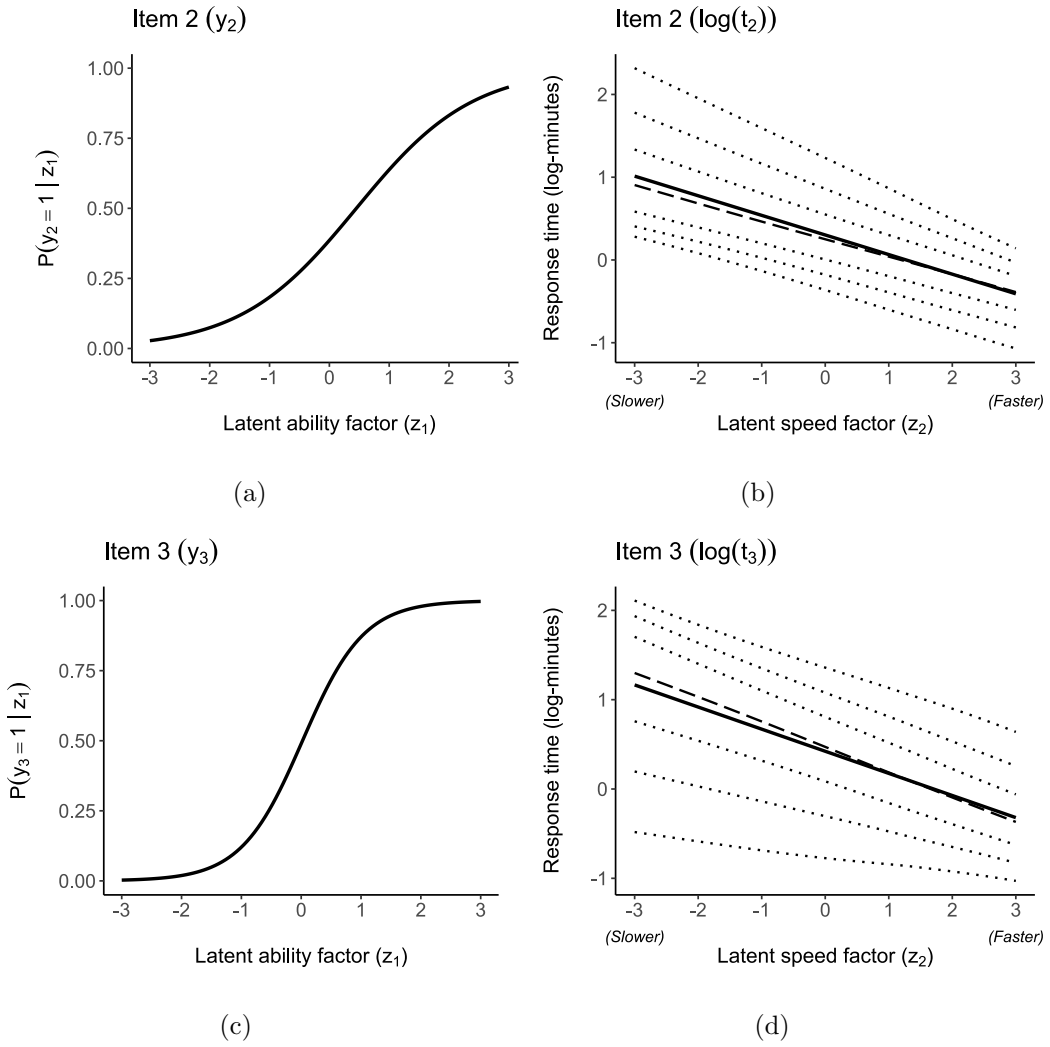


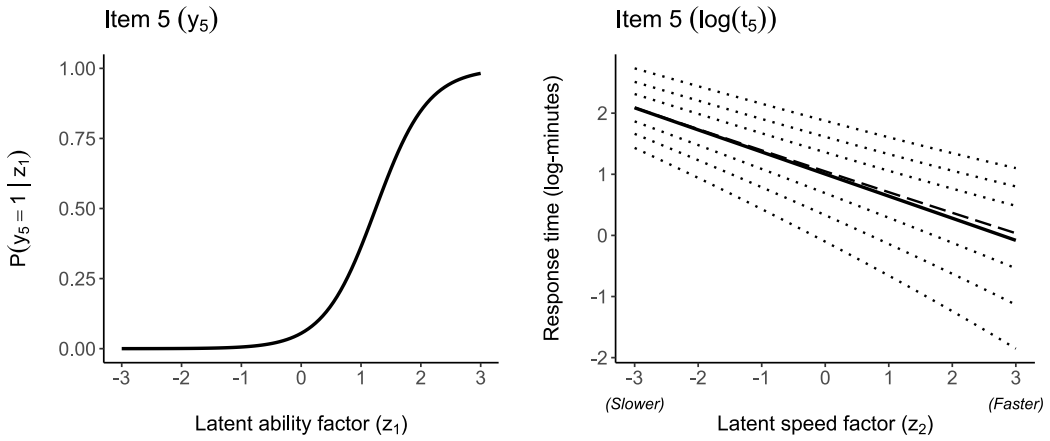
Figure 4: PISA 2018: Fitted conditional expected values (solid line, ——), median (dashed line, - - -), and percentiles (dotted lines, ·····) for IR and log-RT for items 2 and 3.

4.1 Simulation Study I

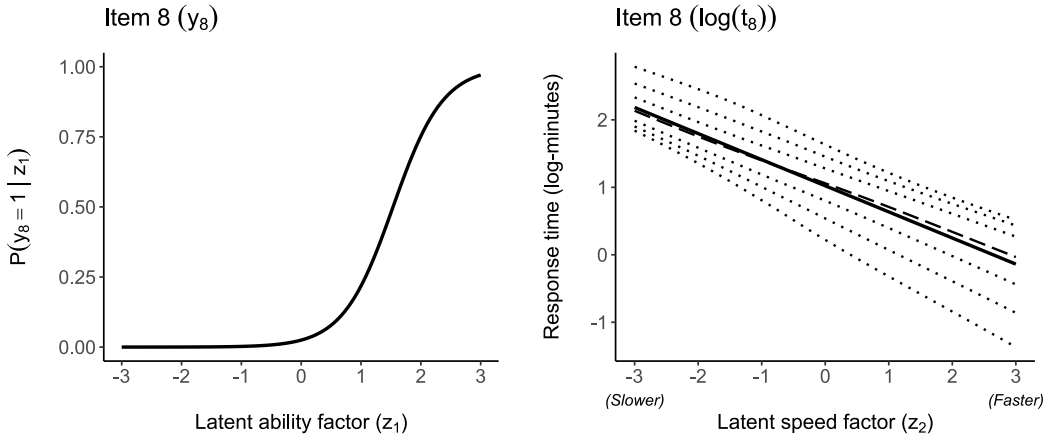
The first simulation study closely resembles the empirical application discussed in Section 3.2. In this study, we examine observed continuous variables within the interval $(0, 1)$, which are assumed to follow a location-scale parametrization of the Beta distribution. Specifically, the heteroscedastic Beta factor model features location and scale parameters that are defined as functions of a single latent variable, i.e., $y_i \mid z \sim \text{Beta}(\mu_i(z), \sigma_i(z))$.

The values for the population parameters in the location parameter equation are selected from two uniform distributions: $\alpha_{i0,\mu} \sim \text{Unif}(-1.5, 1.5)$ and $\alpha_{i1,\mu} \sim \text{Unif}(0.5, 1)$. The signs of the $\alpha_{i1,\mu}$'s are assigned randomly with a probability of 0.5. The parameters for the scale equation are sampled from the uniform distribution $(\alpha_{i0,\sigma}, \alpha_{i1,\mu})^\intercal \sim$

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press



(a) (b)



(c) (d)

Figure 5: PISA 2018: Fitted conditional expected values (solid line, ——), median (dashed line, - - -), and percentiles (dotted lines, ·····) for IR and log-RT for items 5 and 8.

Unif$(0.1, 0.5)$, with the signs of the slopes also assigned randomly. We generate the true parameters in this way to ensure that the conditional densities $f_i(y_i \mid z)$ are uni-modal. Although the Beta distribution allows for bimodal densities for certain combinations of μ_i and σ_i, this is not common in the applications of interest (see Noel (2014) for a uni-dimensional unfolding Beta factor model that handles the bi-modality of the observed variables). The integrals involved in parameter computation were numerically evaluated using a fixed-point Gauss-Hermite rule with 100 quadrature points.

We generated $R = 300$ datasets for each of the 12 conditions. These conditions were created by combining four different sample sizes (200, 500, 1000, and 5000) with three different numbers of observed variables (5, 10, and 20). The quality of the estimated

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

parameters was assessed by the mean squared error (MSE):

$$\text{MSE}(\hat{\alpha}_k) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\alpha}_k^{(r)} - \alpha_k)^2, \quad k = 1, ..., K\,,$$

and the absolute bias (AB):

$$\text{AB}(\hat{\alpha}_k) = \left|\frac{1}{R}\sum_{r=1}^{R}\hat{\alpha}_k^{(r)} - \alpha_k\right|, \quad k = 1, ..., K\,;$$

where $\hat{\alpha}_k^{(r)} \in \hat{\Theta}^{(r)}$ is an arbitrary parameter estimate from the r^{th} replication, and $\alpha_k \in \Theta^*$ is the true value for the model parameter. We report the average MSE (AvMSE) and the average AB (AvAB) separately for the intercepts and slopes in the equations of the location (μ) and scale (σ) parameters indexing the Beta distribution. For completeness, we include box-plots with the simulation results for individual parameters ($p = 10$) in the Online Supplementary Material A5. Similar results hold for other numbers of observed variables.

To evaluate the accuracy of the estimated standard errors (SE) and corresponding confidence intervals, we calculate the average coverage rate across replications for various intercepts and slopes in the location and scale parameters equations. The coverage rate (CR) of the $(1-\alpha)\times 100\%$ confidence interval for a parameter estimate $\hat{\alpha} \in \hat{\Theta}$ is:

$$\text{CR}_\alpha(\hat{\alpha}_k) = \frac{1}{R}\sum_{r=1}^{R}\mathbb{1}(\hat{\text{L}}_k^{(r)} \leq \alpha_k \leq \hat{\text{U}}_k^{(r)}), \quad k = 1, ..., K$$

where $\hat{\text{L}}_k^{(r)} = \hat{\alpha}_k^{(r)} - z_{\alpha/2}\cdot\hat{\text{SE}}(\hat{\alpha}_k^{(r)})$ is the sample-dependent lower bound, $\hat{\text{U}}_k^{(r)} = \hat{\alpha}_k^{(r)} + z_{\alpha/2}\cdot\hat{\text{SE}}(\hat{\alpha}_k^{(r)})$ is the sample-dependent upper bound, and $z_{\alpha/2}$ corresponds to the $(\alpha/2)^{\text{th}}$ quantile of the standard Normal distribution. The customary level of the nominal rate is $\alpha = 0.05$. Coverage rates close to 0.95 indicate a good estimation of the 95% confidence intervals. We report the average CR (AvCR) for intercepts and slopes separately.

Table 5 gives all the results. As expected, the AvMSE and the AvAB tend to decrease as the sample size increases for all values of p. As for the AvCRs, they reach the nominal level as the sample size increases. It is worth noting that estimated standard errors are slightly overestimated for smaller sample sizes, leading to more conservative confidence intervals.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

p	n	Average MSE (AvMSE)				Average AB (AvAB)				Average CR (AvCR)				
		Inter. ($\hat{\alpha}_{i0,\mu}$)	Load. ($\hat{\alpha}_{i1,\mu}$)	Inter. ($\hat{\alpha}_{i0,\sigma}$)	Load. ($\hat{\alpha}_{i1,\sigma}$)	Inter. ($\hat{\alpha}_{i0,\mu}$)	Load. ($\hat{\alpha}_{i1,\mu}$)	Inter. ($\hat{\alpha}_{i0,\sigma}$)	Load. ($\hat{\alpha}_{i1,\sigma}$)	Inter. ($\hat{\alpha}_{i0,\mu}$)	Load. ($\hat{\alpha}_{i1,\mu}$)	Inter. ($\hat{\alpha}_{i0,\sigma}$)	Load. ($\hat{\alpha}_{i1,\sigma}$)	Average CT (mins.)
5	200	0.0105	0.0146	0.0102	0.0081	0.0037	0.0058	0.0240	0.0058	0.9587	0.9720	0.9587	0.9720	0.0415
	500	0.0045	0.0055	0.0039	0.0031	0.0021	0.0050	0.0081	0.0058	0.9507	0.9633	0.9547	0.9660	0.1194
	1000	0.0021	0.0026	0.0018	0.0015	0.0016	0.0050	0.0038	0.0029	0.9553	0.9573	0.9540	0.9620	0.2286
	5000	0.0004	0.0006	0.0004	0.0003	0.0008	0.0044	0.0020	0.0032	0.9580	0.9420	0.9527	0.9473	1.4237
10	200	0.0117	0.0109	0.0079	0.0064	0.0050	0.0066	0.0153	0.0042	0.9630	0.9810	0.9640	0.9917	0.0851
	500	0.0041	0.0042	0.0029	0.0026	0.0028	0.0060	0.0073	0.0036	0.9663	0.9733	0.9583	0.9723	0.2409
	1000	0.0021	0.0021	0.0014	0.0014	0.0012	0.0042	0.0040	0.0042	0.9570	0.9600	0.9590	0.9577	0.4608
	5000	0.0004	0.0004	0.0003	0.0003	0.0009	0.0038	0.0014	0.0037	0.9517	0.9470	0.9520	0.9403	3.0741
20	200	0.0103	0.0094	0.0067	0.0059	0.0060	0.0045	0.0161	0.0050	0.9860	0.9973	0.9853	0.9967	0.1842
	500	0.0042	0.0037	0.0026	0.0024	0.0034	0.0050	0.0056	0.0039	0.9670	0.9790	0.9653	0.9780	0.4937
	1000	0.0020	0.0018	0.0013	0.0012	0.0020	0.0039	0.0040	0.0030	0.9605	0.9707	0.9555	0.9693	1.0251
	5000	0.0004	0.0004	0.0003	0.0003	0.0012	0.0031	0.0010	0.0028	0.9503	0.9452	0.9477	0.9440	6.5482

Table 5: Simulation Study I: Average Mean Squared Error (AvMSE), Average Absolute Bias (AvAB), Average Coverage Rate (AvCR), and Average computation time in minutes (CT) for the MLE of an LVM with Beta distributed observed variables, by test length and sample size. Performance measures are computed for the estimated parameters $\hat{\alpha}_k$ in the location loading matrix ($\hat{\mathrm{A}}_\mu$) and scale loading matrix ($\hat{\mathrm{A}}_\sigma$).

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

4.2 Simulation Study II

The second study resembles the empirical application in Section 3.2. We study the finite sample performance of a confirmatory GLVM-LSS model with two latent variables ($q = 2$) and sixteen observed variables ($p = 16$). The first eight variables are distributed as Bernoulli, conditional on the first factor, $y_i \mid z_1 \sim \text{Bernoulli}(\pi_i(z_1))$ for $i = 1, ..., 8$. The remaining eight variables are distributed Skew-Normal, conditional on the second factor, $y_i \mid z_2 \sim \text{SN}(\mu_i(z_2), \sigma_i(z_2), \nu_i(z_2))$ for $i = 9, ..., 16$. The latent variables follow a multivariate standard Normal distribution, $(z_1, z_2)^\intercal \sim \mathbb{N}(\mathbf{0}, \mathbf{\Phi})$, where $\mathbf{\Phi}$ is a correlation matrix with non-zero off-diagonal entries denoted by $\phi_{12} = \phi_{21} = \phi_{\mathbf{z}}$.

The simulation study considers three sample sizes, $n = \{500, 1000, 3000\}$. The intercepts, slopes, and factor correlation are fixed to the parameter estimates for items 1 to 7 and 9 in Table 4[5]. As before, we compute the AvMSE and AvAB for the estimated parameters and assess the properties of the estimated confidence intervals via the AvCR. The above measures were obtained from $R = 300$ independently simulated datasets. The multidimensional integrals were numerically evaluated using the Gauss-Hermite rule with 35 quadrature points on each latent dimension (a total of 1225 quadrature points).

For better numerical stability, we estimate the model parameters by letting the EM algorithm run for a large number of iterations, using a gradient descent update rule with adaptive learning rate[6]. After 500 iterations, the algorithm switches to the quasi-Newton direct maximization step. Table 6 presents the results for each sample size.

For simplicity, we present the aggregate results for intercepts and factor loadings in each matrix $\hat{\text{A}}_\varphi$, $\varphi \in \boldsymbol{\theta}$. In all cases, the AvMSE and AvAB decrease with sample size, as expected. The results in Table 6 also suggest that the factor correlation $\phi_{\mathbf{z}}$ is consistently estimated. Coverage rates are around nominal levels for medium and large samples, while, as discussed previously, the confidence intervals for the smaller sample size are slightly conservative.

5. Discussion

This paper presents a general framework for latent variable modeling called Generalized Latent Variable Models for Location, Scale, and Shape parameters (GLVM-LSS). In this framework, all the distributional parameters characterizing each observed variable's con-

[5]These eight items were chosen randomly from the nine in the original PISA 2018 application.

[6]The initial learning rate was $\kappa = 0.0001$, and it was halved if the objective function in the EM algorithm did not increase between iterations.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

n	Parameter	Average MSE (AvMSE)	Average AB (AvAB)	Average CR (AvCR)	Average CT (AvCT, mins.)
500	$\hat{\mathrm{A}}_{\pi}$	0.1166	0.0398	0.9631	24.996
	$\hat{\mathrm{A}}_{\mu}$	0.0006	0.0012	0.9613	
	$\hat{\mathrm{A}}_{\sigma}$	0.0017	0.0039	0.9672	
	$\hat{\mathrm{A}}_{\nu}$	0.2033	0.0498	0.9775	
	$\hat{\phi}_{\mathbf{z}}$	0.0036	0.0069	0.9860	
1000	$\hat{\mathrm{A}}_{\pi}$	0.0553	0.0173	0.9599	70.152
	$\hat{\mathrm{A}}_{\mu}$	0.0003	0.0009	0.9597	
	$\hat{\mathrm{A}}_{\sigma}$	0.0008	0.0023	0.9607	
	$\hat{\mathrm{A}}_{\nu}$	0.0775	0.0212	0.9626	
	$\hat{\phi}_{\mathbf{z}}$	0.0017	0.0013	0.9730	
3000	$\hat{\mathrm{A}}_{\pi}$	0.0203	0.0100	0.9496	197.136
	$\hat{\mathrm{A}}_{\mu}$	0.0001	0.0003	0.9575	
	$\hat{\mathrm{A}}_{\sigma}$	0.0003	0.0010	0.9563	
	$\hat{\mathrm{A}}_{\nu}$	0.0209	0.0101	0.9581	
	$\hat{\phi}_{\mathbf{z}}$	0.0005	0.0006	0.9867	

Table 6: Simulation Study II: Average Mean Squared Error (AvMSE), Average Absolute Bias (AvAB), Average Coverage Rate (AvCR), and Average computation time in minutes (CT) for the MLE of a confirmatory GLVM-LSS with Bernoulli and Skew-Normal distributed observed variables, by sample size and type of parameter. Performance measures are computed for the estimated parameters in the loading matrix for the Bernoulli items ($\hat{\mathrm{A}}_{\pi}$); the loading matrices for the location ($\hat{\mathrm{A}}_{\mu}$), scale ($\hat{\mathrm{A}}_{\sigma}$), and shape ($\hat{\mathrm{A}}_{\nu}$) parameters for the Skew-Normal items; and the correlation between the latent variables ($\hat{\phi}_{\mathbf{z}}$).

ditional distribution are modeled as functions of linear combinations of the latent variables. In this respect, the (conditional) mean and higher-order (conditional) moments of the observed variables -expressed in terms of the corresponding location, scale, and shape parameters- are also considered functions of the latent variables. The GLVM-LSS offers a wide range of possibilities for modeling complex multivariate datasets by allowing the modeling of data displaying heteroscedasticity, excess skewness, excess kurtosis, zero/one/maximum value inflation, heaping, truncation, or censoring. Model parameters are estimated via full-information maximum likelihood. We demonstrate the effectiveness of our framework by presenting two GLVM-LSS applications using real-world data in public opinion research and educational testing. Our proposed method is implemented in the `R` package `glvmlss`, available online at https://github.com/ccardehu/glvmlss.

The GLVM-LSS framework has numerous potential applications in empirical research

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

areas where modeling higher-order moments of observed variables as functions of latent variables is relevant. For instance, in ecological momentary assessment designs, researchers are interested in individuals' emotional response variability, in addition to the deviations from their baseline mood (Hedeker et al., 2006, 2008, 2012, Wang et al., 2012). While these models include random effects, incorporating a latent variable specification could enrich the analysis. Another example is item quality control in educational testing (Hessen and Dolan, 2009). Heteroscedastic items often exhibit low discrimination power, which can reduce test accuracy if not revised or removed. From a dimensionality reduction perspective, it is desirable to preserve as much information as possible and to model the essential aspects of the observed data. Modeling the entire (conditional) distribution can result in better data recovery than simply modeling the (conditional) mean (Shen and Meinshausen, 2024).

While the GLVM-LSS is a flexible tool for modeling multivariate data with latent variables, there are still opportunities for improvement in future research. Currently, the model's implementation in the `glvmlss` package computes numerical integrals in the MML estimation algorithm and factor scoring procedure using a fixed-point Gaussian-Hermite quadrature rule. This approach limits the number of latent variables that can be included in the model without encountering computational bottlenecks. Future updates of the `glvmlss` package will aim to address this limitation by incorporating either adaptive Gaussian-Hermite quadrature rules (Rabe-Hesketh et al., 2005) or stochastic approximation methods (Cai, 2010, Zhang and Chen, 2022). These alternatives for numerical integration have been shown to produce fast and accurate solutions and thus should be explored further in the context of the GLVM-LSS.

As discussed earlier, one potential extension of the GLVM-LSS framework is to incorporate observed covariates into both the measurement and structural equations. This addition could enhance the framework's usefulness in latent regression contexts and aid in testing for measurement invariance or differential item functioning. Also, many datasets in the social sciences suffer from non-ignorable missingness, and thus, one can extend the GLVM-LSS and implement methods developed, for example, in O'Muircheartaigh and Moustaki (1999).

Moreover, increasing the flexibility to model distributional parameters also raises the number of parameters that need to be estimated, which can lead to computational and interpretational challenges. A regularized estimation approach for the GLVM-LSS can be developed to mitigate these issues. This method aims to produce more sparse and interpretable factor loading solutions while also facilitating model selection through the appropriate choice of the regularization parameter (e.g., Geminiani et al., 2021, Cárdenas-

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Hurtado, 2023).

Another extension of the GLVM-LSS framework involves relaxing the linearity assumption in the specification of the (linear) predictor $\eta_{i,\varphi}(\mathbf{z})$. Following the generalized additive model specification in the GAMLSS regression framework, for each distributional parameter $\varphi_i \in \boldsymbol{\theta}_i$ we could model the relationship between the observed and latent variables using splines (Ramsay and Abrahamowicz, 1989, de Boor, 2001). This approach extends previous research in non-linear latent variable models using polynomials (McDonald, 1962, 1967, Yalcin and Amemiya, 2001, Rizopoulos and Moustaki, 2008). It also has connections with existing models in the LVM literature, such as the unidimensional semi-parametric IRT models (e.g., Ramsay and Winsberg, 1991, Rossi et al., 2002, Johnson, 2007, Falk and Cai, 2016b,a).

Acknowledgments We thank the Editor, Associate Editor, and three anonymous referees for their careful review and valuable comments.

Funding Statement This research received no specific grant funding form any funding agency, commercial or not-for-profit sectors.

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

Data Availability Statement The associated `glvmlss` package and replication files are available online at https://github.com/ccardehu/glvmlss.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

References

Abelson, R. P., Kinder, D. R., Peters, M. D., and Fiske, S. T. (1982). Affective and Semantic Components in Political Person Perception. *Journal of Personality and Social Psychology*, 42(4):619–630.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Allman, E., Matias, C., and Rhodes, J. A. (2009). Identifiability of Parameters in Latent Structure Models with many Observed Variables. *The Annals of Statistics*, 37(6A):3099–3132.

Anderson, T. W. and Rubin, H. (1956). Statistical Inference in Factor Analysis. In Neyman, J., editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1955*, volume V, pages 111–150. University of California Press.

Asparouhov, T. and Muthén, B. (2016). Structural Equation Models and Mixture Models With Continuous Nonnormal Skewed Distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4):1–19.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2):171–178.

Azzalini, A. (2005). The Skew-normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, 35(2):159–188.

Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 61(3):579–602.

Balamuta, J. J. and Eddelbuettel, D. (2018). *RcppEnsmallen: Header-Only C++ Mathematical Optimization Library for 'Armadillo'*.

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley Series in Probability and Statistics. New York, NY, US: John Wiley & Sons, Ltd, 3rd edition.

Bock, R. D. and Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM algorithm. *Psychometrika*, 46(4):443–459.

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York, NY, US: John Wiley & Sons, Ltd., 1st edition.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Bollen, K. A. (1996). A limited-information estimator for LISREL models with or without heteroscedastic errors. In Marcoulides, G. A. and Schumacker, R. E., editors, *Advanced Structural Equation Modeling: Issues and Techniques*, pages 227–241. Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.

Bolsinova, M. and Molenaar, D. (2018). Modeling Nonlinear Conditional Dependence Between Response Time and Accuracy. *Frontiers in Psychology*, 9(1525):1–12.

Bolsinova, M., Tijmstra, J., and Molenaar, D. (2017). Response moderation models for conditional dependence between response time and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 70(2):257–279.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1):62–83.

Cai, L. (2010). High-dimensional Exploratory Item Factor Analysis by a Metropolis–Hastings Robbins–Monro Algorithm. *Psychometrika*, 75(1):33–57.

Cárdenas-Hurtado, C. A. (2023). *Generalised Latent Variable Models for Location, Scale, and Shape parameters*. PhD thesis, London School of Economics and Political Science.

Choirat, C. and Seri, R. (2012). Estimation in Discrete Parameter Models. *Statistical Science*, 27(2):278–293.

Davidov, E., Schmidt, P., Billiet, J., and Meuleman, B., editors (2018). *Cross-Cultural Analysis: Methods and Applications*. European Association of Methodology Series. New York, NY, US: Routledge, Taylor & Francis group, 2nd edition.

De Boeck, P. and Jeon, M. (2019). An Overview of Models for Response Times and Processes in Cognitive Tests. *Frontiers in Psychology*, 10(102):1–11.

de Boor, C. (2001). *A Practical Guide to Splines (Revised Edition)*, volume 27 of *Springer Series in Applied Mathematical Sciences*. New York, NY, US: Springer-Verlag, 2nd edition.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–38.

Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Russell, N., Ucar, I., Bates, D., and Chambers, J. (2023a). *Rcpp: Seamless R and C++ Integration*. R package version 1.0.11.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Eddelbuettel, D., Francois, R., Bates, D., Ni, B., and Sanderson, C. (2023b). *RcppArmadillo: 'Rcpp' Integration for the 'Armadillo' Templated Linear Algebra Library*. R package version 0.12.6.4.0.

Falk, C. F. and Cai, L. (2016a). Maximum Marginal Likelihood Estimation of a Monotonic Polynomial Generalized Partial Credit Model with Applications to Multiple Group Analysis. *Psychometrika*, 81(2):434–460.

Falk, C. F. and Cai, L. (2016b). Semiparametric Item Response Functions in the Context of Guessing. *Journal of Educational Measurement*, 53(2):229–247.

Geminiani, E., Marra, G., and Moustaki, I. (2021). Single- and Multiple-Group Penalized Factor Analysis: A Trust-Region Algorithm Approach with Integrated Automatic Multiple Tuning Parameter Selection. *Psychometrika*, 86(1):65–95.

Greene, W. H. (2003). *Econometric Analysis*. Upper Saddle River, NJ, US: Prentice-Hall Inc., 5th edition.

Gu, Y. and Xu, G. (2020). Partial Identifiability of Restricted Latent Class Models. *The Annals of Statistics*, 48(4):2082–2107.

Guth, J. L. (2019). Are White Evangelicals Populists? The View from the 2016 American National Election Study. *The Review of Faith and International Affairs*, 17(3):20–35.

Hammersley, J. M. (1950). On Estimating Restricted Parameters. *Journal of the Royal Statistical Society: Series B (Methodological)*, 12(2):192–229.

Hedeker, D., Berbaum, M., and Mermelstein, R. (2006). Location-Scale Models for Multilevel Ordinal Data: Between- and Within-Subjects Variance Modeling. *Journal of Probability and Statistical Science*, 4(1):1–20.

Hedeker, D., Demirtas, H., and Mermelstein, R. (2009). A mixed ordinal location scale model for analysis of ecological momentary assessment (EMA) data. *Statistics and its Interface*, 2(4):391–401.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Wiley Series in Probability and Statistics. New York, NY, US: John Wiley & Sons, Ltd., 1st edition.

Hedeker, D., Mermelstein, R., and Demirtas, H. (2008). An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data. *Biometrics*, 64(2):627–634.

Hedeker, D., Mermelstein, R., and Demirtas, H. (2012). Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine*, 31(27):3328–3336.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Hedeker, D., Mermelstein, R. J., Demirtas, H., and Berbaum, M. L. (2016). A mixed-effects location-scale model for ordinal questionnaire data. *Health Services and Outcomes Reseach Methodology*, 16(3):117–131.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8:1–19.

Hessen, D. J. and Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology*, 62(1):57–77.

Jennrich, R. I. (2004). Rotation to Simple Loadings Using Component Loss Functions: The Orthogonal Case. *Psychometrika*, 69(2):257–273.

Jennrich, R. I. (2006). Rotation to Simple Loadings Using Component Loss Functions: The Oblique Case. *Psychometrika*, 71(1):173–191.

Johnson, M. S. (2007). Modeling dichotomous item responses with free-knot splines. *Computational Statistics & Data Analysis*, 51:4178–4192.

Jöreskog, K. G. and Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, 36(3):347–387.

Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in germany. *The Annals of Applied Statistics*, 9(2):1024–1052.

Krasa, S. and Polborn, M. (2014). Policy Divergence and Voter Polarization in a Structural Model of Elections. *Journal of Law and Economics*, 57(1):31–76.

Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*, 33(2):188–229.

Lai, K. (2018). Estimating Standardized SEM Parameters Given Nonnormal Data and Incorrect Model: Methods and Comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4):600–620.

Lei, M. and Lomax, R. G. (2005). The Effect of Varying Degrees of Nonnormality in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1):1–27.

Lewin-Koh, S.-C. and Amemiya, Y. (2003). Heteroscedastic factor analysis. *Biometrika*, 90(1):85–97.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Liu, M. and Lin, T. I. (2015). Skew-normal factor analysis models with incomplete data. *Journal of Applied Statistics*, 42(4):789–805.

Liu, X., Wallin, G., Chen, Y., and Moustaki, I. (2023). Rotation to Sparse Loadings using L^p Losses and Related Inference Problems. *Psychometrika*, 88(2):527–553.

Loeys, T., Rosseel, Y., and Baten, K. (2011). A Joint Modeling Approach for Reaction Time and Accuracy in Psycholinguistic Experiments. *Psychometrika*, 76(3):487–503.

Louis, T. A. (1982). Finding the Observed Information Matrix When Using the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233.

Magnus, B. E. and Thissen, D. (2017). Item Response Modeling of Multivariate Count Data With Zero Inflation, Maximum Inflation, and Heaping. *Journal of Educational and Behavioral Statistics*, 42(5):531–558.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, 27(4):397–415.

McDonald, R. P. (1967). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, 32(1):77–112.

McDonald, R. P. and Krane, W. R. (1977). A note on local identifiability and degrees of freedom in the asymptotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology*, 30(2):198–203.

Molenaar, D. (2015). Heteroscedastic Latent Trait Models for Dichotomous Data. *Psychometrika*, 80(3):625–644.

Molenaar, D., Cúri, M., and Bazán, J. L. (2022). Zero and One Inflated Item Response Theory Models for Bounded Continuous Data. *Journal of Educational and Behavioral Statistics*, 47(6):693–735.

Molenaar, D., Dolan, C. V., and de Boeck, P. (2012). The Heteroscedastic Graded Response Model with a Skewed Latent Trait: Testing Statistical and Substantive Hypotheses Related to Skewed Item Category Functions. *Psychometrika*, 77(3):455–478.

Molenaar, D., Dolan, C. V., and Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, 63(2):293–317.

Molenaar, D., Tuerlinckx, F., and van der Maas, H. L. J. (2015). A Bivariate Generalized Linear Item Response Theory Modeling Framework to the Analysis of Responses and Response Times. *Multivariate Behavioral Research*, 50(1):56–74.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Montanari, A. and Viroli, C. (2010). A skew-normal factor model for the analysis of student satisfaction towards university courses. *Journal of Applied Statistics*, 37(3):473–487.

Moustaki, I. and Knott, M. (2000). Generalized Latent Trait Models. *Psychometrika*, 65(3):391–411.

Moustaki, I. and Steele, F. (2005). Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in Bangladesh. *Statistical Modelling*, 5(4):327–342.

Moustaki, I. and Victoria-Feser, M. P. (2006). Bounded-Influence Robust Estimation in Generalized Linear Latent Variable Models. *Journal of the American Statistical Association*, 101(474):644–653.

Niku, J., Warton, D. I., Hui, F. K. C., and Taskinen, S. (2017). Generalized Linear Latent Variable Models for Multivariate Count and Biomass Data in Ecology. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(4):498–522.

Nishii, R. (1984). Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression. *The Annals of Statistics*, 12(2):758–765.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. New York, NY, US: Springer-Verlag, 2nd edition.

Noel, Y. (2014). A Beta Unfolding Model for Continuous Bounded Responses. *Psychometrika*, 79(4):647–674.

Noel, Y. and Dauvier, B. (2007). A Beta Item Response Model for Continuous Bounded Responses. *Applied Psychological Measurement*, 31(1):47–73.

Ondish, P. and Stern, C. (2018). Liberals Possess More National Consensus on Political Attitudes in the United States: An Examination Across 40 Years. *Social Psychological and Personality Science*, 9(8):935–943.

O'Muircheartaigh, C. and Moustaki, I. (1999). Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(2):177–194.

R Core Team (2022). *R: A Language and Environment for Statistical Computing, version 4.2.2*. R Foundation for Statistical Computing, Vienna, Austria.

Rabe-Hesketh, S., Skrondal, A., and Pickels, A. (2004). Generalized Multilevel Structural Equation Modelling. *Psychometrika*, 69(2):167–190.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Rabe-Hesketh, S., Skrondal, A., and Pickels, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128(2):301–323.

Ramsay, J. O. and Abrahamowicz, M. (1989). Binomial Regression With Monotone Splines: A Psychometric Application. *Journal of the American Statistical Association*, 84(408):906–915.

Ramsay, J. O. and Winsberg, S. (1991). Maximum marginal likelihood estimation for semiparametric item analysis. *Psychometrika*, 56(3):365–379.

Revuelta, J., Hidalgo, B., and Alcazar-Córcolesa, M. A. (2022). Bayesian Estimation and Testing of a Beta Factor Model for Bounded Continuous Variables. *Multivariate Behavioral Research*, 57(1):1–22.

Rigby, R. A. and Stasinopoulos, M. D. (2005). Generalized additive models for location, shape and scale. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., and De Bastiani, F. (2020). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. Chapman & Hall/CRC The R Series. Boca Ratón, FL, US: Chapman & Hall / CRC.

Rizopoulos, D. and Moustaki, I. (2008). Generalized latent variable models with non-linear effects. *British Journal of Mathematical and Statistical Psychology*, 61(2):415–438.

Rossi, N., Wang, X., and Ramsay, J. O. (2002). Nonparametric Item Response Function Estimates with the EM Algorithm. *Journal of Educational and Behavioral Statistics*, 27(3):291–317.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica*, 39(3):577–591.

Schilling, S. and Bock, R. D. (2005). High-dimensional Maximum Marginal Likelihood Item Factor Analysis by Adaptive Quadrature. *Psychometrika*, 70(3):533–555.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Shao, J. (1997). An Asymptotic Theory for Linera Model Selection. *Statistica Sinica*, 7(2):221–242.

Shen, X. and Meinshausen, N. (2024). Distributional Principal Autoencoders. https://arxiv.org/abs/2404.13649.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: multilevel, longitudinal, and structural equation models*. Interdisciplinary Statistics. Boca Ratón, FL, US: Chapman & Hall, CRC.

Smithson, M. and Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables. *Psychological Methods*, 11(1):54–71.

Stasinopoulos, M. D., , Kneib, T., Klein, N., Mayr, A., and Heller, G. Z. (2024). *Generalized Additive Models for Location, Scale, and Shape: A Distributional Regression Approach with Applications*. Number 56 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press, 1st edition.

Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F. (2017). *Flexible regression and smoothing: Using GAMLSS in R*. Chapman & Hall/CRC The R Series. Boca Ratón, FL, US: Chapman & Hall / CRC.

Umlauf, N., Klein, N., and Zeileis, A. (2018). BAMLSS: Bayesian Additive Models for Location, Scale, and Shape (and Beyond). *Journal of Computational and Graphical Statistics*, 27(3):612–627.

van der Linden, W. J. (2006). A Lognormal Model for Response Times on Test Items. *Journal of Educational and Behavioral Statistics*, 31(2):181–204.

van der Linden, W. J. (2007). A Hierarchical Framework for Modeling Speed and Accuracy on Test Items. *Psychometrika*, 72(3):287–308.

van der Linden, W. J. (2008). Using Response Times for Item Selection in Adaptive Testing. *Journal of Educational and Behavioral Statistics*, 33(1):5–20.

van der Linden, W. J. (2009). Conceptual Issues in Response-Time Modeling. *Journal of Educational Measurement*, 46(3):247–272.

van der Linden, W. J. and Guo, F. (2008). Bayesian Procedures for Identifying Aberrant Response-Time Patterns in Adaptive Testing. *Psychometrika*, 73(3):365–384.

van der Linden, W. J., Klein Entink, R. H., and Fox, J.-P. (2010). IRT Parameter Estimation With Response Times as Collateral Information. *Applied Psychological Measurement*, 34(5):327–347.

Van Zandt, T. (2002). Analysis of response time distributions. In Wixted, J. T. and Pashler, H., editors, *Stevens' Handbook of Experimental Psychology*, volume 4 of *Methodology in Experimental Psychology*, page 461–516. New York, NY, US: John Wiley & Sons, Ltd, 3rd edition.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press

Verkuilen, J. and Smithson, M. (2012). Mixed and Mixture Regression Models for Continuous Bounded Responses Using the Beta Distribution. *Journal of Educational and Behavioral Statistics*, 37(1):82–113.

Wall, M. M., Park, J. Y., and Moustaki, I. (2015). IRT Modeling in the Presence of Zero-Inflation With Application to Psychiatric Disorder Severity. *Applied Psychological Measurement*, 39(8):583–597.

Wang, L. (2010). IRT-ZIP Modeling for Multivariate Zero-Inflated Count Data. *Journal of Educational and Behavioral Statistics*, 35(6):671–692.

Wang, L., Hamaker, E., and Bergeman, C. S. (2012). Investigating Inter-Individual Differences in Short-Term Intra-Individual Variability. *Psychological Methods*, 17(4):567–581.

Yalcin, I. and Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, 16(3):275–294.

Zhang, S. and Chen, Y. (2022). Computation for Latent Variable Model Estimation: A Unified Stochastic Proximal Framework. *Psychometrika*, 87(4):1473–1502.

Zimmerman, M. E. (2011). Speed-accuracy tradeoff. In Kreutzer, J. S., DeLuca, J., and Caplan, B., editors, *Encyclopedia of Clinical Neuropsychology*, pages 2344–2344. New York, NY, US: Springer-Verlag.

https://doi.org/10.1017/psy.2025.7 Published online by Cambridge University Press