Generalised Latent Variable Models for Location, Scale and Shape parameters

Camilo A. Cárdenas-Hurtado

Joint work with
Irini Moustaki & Yunxiao Chen (LSE)
Giampiero Marra (UCL)

M3 Conference

June 27th 2023

Background: Latent Variable Models (LVM)

- Observed variables (items): $\mathbf{y} = (y_1, ..., y_p)^\intercal \in \mathbb{R}^p$,
- Latent variables (factors): $\mathbf{z} = (z_1, ..., z_q)^\intercal \in \mathbb{R}^q$, with $q \ll p$.

The marginal distribution:

$$f(\mathbf{y}; \Theta) = \int_{\mathbb{R}^q} \prod_{i=1}^{p} f_i(y_i \mid \mathbf{z}; \Theta_{y_i}) \; p(\mathbf{z}; \Theta_z) \; d\mathbf{z}$$

Assumptions:

- Parametric model: $\Theta^\intercal = (\Theta_y^\intercal, \Theta_z^\intercal)$,
- Conditional independence,

Background: Latent Variable Models (LVM)

- Observed variables (items): $\mathbf{y} = (y_1, ..., y_p)^\intercal \in \mathbb{R}^p$,
- Latent variables (factors): $\mathbf{z} = (z_1, ..., z_q)^\intercal \in \mathbb{R}^q$, with $q \ll p$.

The marginal distribution:

$$f(\mathbf{y}; \Theta) = \int_{\mathbb{R}^q} \prod_{i=1}^{p} \exp\left\{ \frac{y_i \zeta_i(\mathbf{z}) - b_i(\zeta_i(\mathbf{z}))}{\phi_i} + c_i(y_i; \phi_i) \right\} p(\mathbf{z}; \mathbf{\Phi}) \, \mathbf{dz}$$

Assumptions (GLLVM, Skrondal and Rabe-Hesketh, 2004; Bartholomew et al., 2011):

- Parametric model: $\Theta^\intercal = (\Theta_y^\intercal, \Theta_z^\intercal)$,
- Conditional independence,
- Distributions: $y_i \,|\, \mathbf{z} \sim \text{EF}(\zeta_i(\mathbf{z}), \phi_i)$ and $\mathbf{z} \sim \mathbb{N}(\mathbf{0}, \mathbf{\Phi})$.

Background: Generalised Linear LVM (GLLVM)

$$f(\mathbf{y};\Theta) = \int_{\mathbb{R}^q} \prod_{i=1}^{p} \exp\left\{ \frac{y_i\zeta_i(\mathbf{z}) - b_i(\zeta_i(\mathbf{z}))}{\phi_i} + c_i(y_i;\phi_i) \right\} p(\mathbf{z};\boldsymbol{\Phi})\ \mathrm{d}\mathbf{z}$$

- System of GLMs (McCullagh and Nelder, 1989) with latent covariates:
  - Focus on (conditional) mean: $\mu_i := \mathbb{E}(y_i \mid \mathbf{z}) = \partial b_i / \partial \zeta_i$,
  - Linear predictor: $\upsilon_i(\mu_i) = \alpha_{i0} + \boldsymbol{\alpha}_{i1}^{\mathsf{T}} \mathbf{z}$
  - Matrix notation: $\upsilon(\boldsymbol{\mu}) = \boldsymbol{\alpha}_0 + \mathrm{A}\mathbf{z}$
- Model parameters: $\Theta^{\mathsf{T}} = (\boldsymbol{\alpha}_0{}^{\mathsf{T}}, \mathrm{vec}(\mathrm{A})^{\mathsf{T}}, \boldsymbol{\phi}^{\mathsf{T}}, \mathrm{vech}(\Phi)^{\mathsf{T}})$
- Rotational indeterminacy: q^2 restrictions on Θ (EFA/CFA).

Background: Generalised Linear LVM (GLLVM)

$$\ell(\Theta;\mathbf{y}) = \sum_{m=1}^{n} \log \left[\int_{\mathbb{R}^q} \prod_{i=1}^{p} f_i(y_{im} \mid \mathbf{z}; \alpha_{i0}, \boldsymbol{\alpha}_i, \phi_i)\; p(\mathbf{z}; \Phi)\; \mathbf{dz} \right]$$

- System of GLMs (McCullagh and Nelder, 1989) with latent covariates:
  - Focus on (conditional) mean: $\mu_i := \mathbb{E}(y_i \mid \mathbf{z}) = \partial b_i / \partial \zeta_i$,
  - Linear predictor: $\upsilon_i(\mu_i) = \alpha_{i0} + \boldsymbol{\alpha}_{i1}^{\mathsf{T}}\mathbf{z}$
  - Matrix notation: $\upsilon(\boldsymbol{\mu}) = \boldsymbol{\alpha}_0 + \mathrm{A}\mathbf{z}$
- Model parameters: $\Theta^{\mathsf{T}} = (\boldsymbol{\alpha}_0{}^{\mathsf{T}}, \mathrm{vec}(\mathrm{A})^{\mathsf{T}}, \boldsymbol{\phi}^{\mathsf{T}}, \mathrm{vech}(\Phi)^{\mathsf{T}})$
- Rotational indeterminacy: q^2 restrictions on Θ (EFA/CFA).

$$\hat{\Theta} = \arg\max_{\Theta \in \Xi} \ell(\Theta;\mathbf{y})$$

Motivation: Response times (educational testing research)

- Interest in response times (van der Linden, 2007, 2009).
- Can we assume $\log(t_i) \,|\, \mathbf{z} \sim \mathbb{N}(\mu_i(\mathbf{z}), \sigma_i^2)$? (⚠)

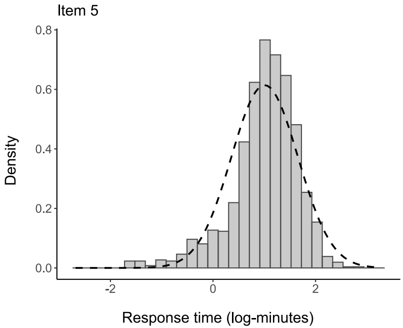


Figure: Empirical vs. marginal distribution assuming $\log(t_i) \,|\, \mathbf{z} \sim \mathbb{N}(\mu_i(\mathbf{z}), \sigma_i^2)$.

Motivation: Thermometer ratings (public opinion research)

- Q: "*From 0 (cold) to 100 (hot), how would you rate _____ ?*"
- Can we assume $y_i \mid \mathbf{z} \sim \mathbb{N}(\mu_i(\mathbf{z}), \sigma_i^2)$? (⚠)

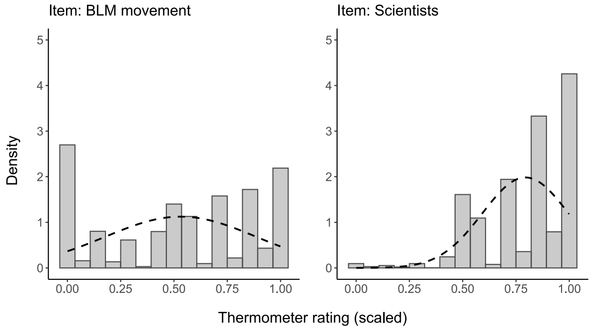


Figure: Empirical vs. marginal distributions assuming $y_i \mid \mathbf{z} \sim \mathbb{N}(\mu_i(\mathbf{z}), \sigma_i^2)$.

Motivation: Limitations of $f(\mathbf{y} \mid \mathbf{z})$ in the GLLVM

A call for modelling the conditional distribution $f_i(y_i \mid \mathbf{z})$, as opposed to only the conditional mean $\mu_i(\mathbf{z})$:

- Going beyond $y_i \mid \mathbf{z} \sim$ EF,
  - e.g, items in $(0, 1)$, zero inflation, heaping (rounding).
- Substantive interest in higher order moments (variance, skewness, kurtosis),
  - e.g., Ability differentiation, Ecological momentary assessments (EMA).
- Item quality and control,
- Better prediction, outliers, etc.

Outline

Outline

Generalised LVM for Location, Scale and Shape (GLVM-LSS)

- We propose an umbrella class of LVM to model the conditional distribution $f_i(y_i \mid \mathbf{z})$.
- System of GAMLSS regressions (Rigby and Stasinopoulos, 2005) with latent covariates:
  - $f_i(y_i \mid \mathbf{z}; \theta_i(\mathbf{z}))$ is indexed by $\theta_i(\mathbf{z}) = (\mu_i, \sigma_i, \nu_i, \tau_i)^\intercal$, a vector of *location* (μ_i), *scale* (σ_i), and *shape* (ν_i, τ_i) parameters, modelled as linear functions of $\mathbf{z}$.
  - Mean, variance, skewness, kurtosis: functions of these distributional parameters.
  - Linear predictor (for $\varphi_i \in \boldsymbol{\theta}_i$): $\upsilon_{i,\varphi}(\varphi_i) = \alpha_{i0,\varphi} + \boldsymbol{\alpha}_{i1,\varphi}^\intercal \mathbf{z}$
  - Matrix notation (for $\varphi \in \boldsymbol{\theta}$): $\upsilon_\varphi(\boldsymbol{\varphi}) = \boldsymbol{\alpha}_{0,\varphi} + \mathrm{A}_\varphi \mathbf{z}$
  - Matrix notation (all): $\upsilon(\boldsymbol{\theta}) = \boldsymbol{\alpha}_0 + \mathrm{A}\mathbf{z}$
- Model parameters: $\Theta^\intercal = (\boldsymbol{\alpha}_0^\intercal, \mathrm{vec}(\mathrm{A})^\intercal, \mathrm{vech}(\Phi)^\intercal)$.

GLVM-LSS: Some examples (1)

- Heteroscedastic Normal linear factor model (Hessen and Dolan, 2009).
- Items: $y_i \,|\, \mathbf{z} \sim \mathbb{N}(\mu_i(\mathbf{z}), \sigma_i^2(\mathbf{z}))$.
- Location ($\mu_i \in \mathbb{R}$) and scale ($\sigma_i \in \mathbb{R}^+$) parameters.
- Measurement equations:

$$\mu_i(\mathbf{z}) = \alpha_{i0,\mu} + \boldsymbol{\alpha}_{i1,\mu}^{\mathsf{T}}\mathbf{z}$$
$$\log(\sigma_i(\mathbf{z})) = \alpha_{i0,\sigma} + \boldsymbol{\alpha}_{i1,\sigma}^{\mathsf{T}}\mathbf{z}$$

- Used for: testing ‘ability differentiation’, item quality control.

Update: Response times (educational testing research)

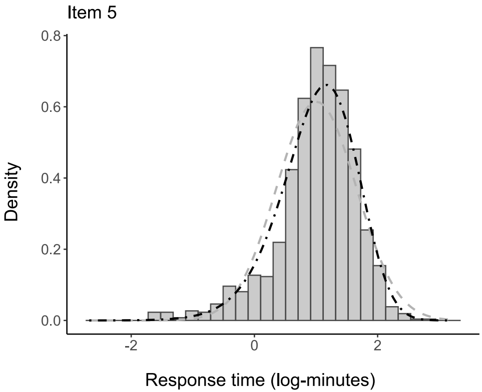


Figure: Homoscedastic, $\mathbb{N}(\mu_i(z_2), \sigma_i^2)$ (– – –) vs. Heteroscedastic, $\mathbb{N}(\mu_i(z_2), \sigma_i^2(z_2))$ (–·–·) model.

GLVM-LSS: Some examples (2)

- Skew-Normal factor model (Montanari and Viroli, 2010; Liu and Lin, 2015; Asparouhov and Muthén, 2016).
- Re-parametrisation: $y_i \,|\, \mathbf{z} \sim \mathsf{SN}(\mu_i(\mathbf{z}), \sigma_i^2(\mathbf{z}), \nu_i(\mathbf{z}))$.
- Location ($\mu_i \in \mathbb{R}$), scale ($\sigma_i \in \mathbb{R}^+$), and shape ($\nu_i \in (0,1)$) parameters.
- Measurement equations:

$$\begin{aligned} \mu_i(\mathbf{z}) &= \alpha_{i0,\mu} + \boldsymbol{\alpha}_{i1,\mu}^\mathsf{T}\mathbf{z} \\ \log(\sigma_i(\mathbf{z})) &= \alpha_{i0,\sigma} + \boldsymbol{\alpha}_{i1,\sigma}^\mathsf{T}\mathbf{z} \\ \mathsf{logit}(\nu_i(\mathbf{z})) &= \alpha_{i0,\nu} + \boldsymbol{\alpha}_{i1,\nu}^\mathsf{T}\mathbf{z} \end{aligned}$$

Update: Response times (educational testing research)

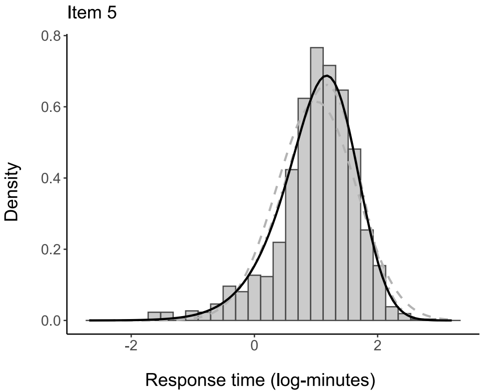


Figure: $\mathbb{N}(\mu_i(z_2), \sigma_i^2)$ (– – –) vs. $\mathbb{N}(\mu_i(z_2), \sigma_i^2(z_2))$ (–·–·) vs. SN$(\mu_i(z_2), \sigma_i^2(z_2), \nu_i(z_2))$ (——).

GLVM-LSS: Some examples (3)

- Zero-Inflated Poisson factor model (Wang, 2010; Wall et al., 2015; Magnus and Thissen, 2017).
- Items: $y_i \,|\, \mathbf{z} \sim \text{ZIP}(\lambda_i(\mathbf{z}), \pi_i(\mathbf{z}))$:

$$f_i(y_i \,|\, \mathbf{z}; \boldsymbol{\theta}_i) = \begin{cases} \pi_i + (1-\pi_i)\cdot e^{-\lambda_i}, & \text{if } y_i = 0 \\ (1-\pi_i)\cdot \dfrac{\lambda_i^{y_i}\cdot e^{-\lambda_i}}{y_i!}, & \text{if } y_i > 0 \end{cases}$$

- Measurement equations:

$$\log(\lambda_i(\mathbf{z})) = \alpha_{i0,\lambda} + \boldsymbol{\alpha}_{i1,\lambda}^{\mathsf{T}}\mathbf{z}$$
$$\text{logit}(\pi_i(\mathbf{z})) = \alpha_{i0,\pi} + \boldsymbol{\alpha}_{i1,\pi}^{\mathsf{T}}\mathbf{z}$$

GLVM-LSS: Some examples (4)

- (Heteroscedastic) Beta factor model (Noel and Dauvier, 2007; Verkuilen and Smithson, 2012; Revuelta et al., 2022).
- Re-parametrisation: $y_i \mid \mathbf{z} \sim \text{Beta}(\mu_i(\mathbf{z}), \sigma_i(\mathbf{z}))$.
- Location ($\mu_i \in (0,1)$), scale ($\sigma_i \in (0,1)$) parameters.
- $\mathbb{E}(y_i \mid \mathbf{z}) = \mu_i$ and $\text{Var}(y_i \mid \mathbf{z}) = \sigma_i^2 \mu_i (1 - \mu_i)$.
- Measurement equations:

$$\text{logit}(\mu_i(\mathbf{z})) = \alpha_{i0,\mu} + \boldsymbol{\alpha}_{i1,\mu}^{\intercal}\mathbf{z}$$
$$\text{logit}(\sigma_i(\mathbf{z})) = \alpha_{i0,\sigma} + \boldsymbol{\alpha}_{i1,\sigma}^{\intercal}\mathbf{z}$$

Update: Thermometer ratings (public opinion research)

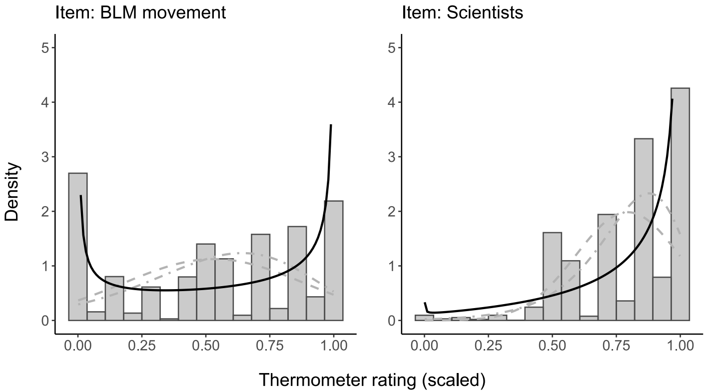


Figure: $\mathbb{N}(\mu_i(z_1), \sigma_i^2)$ (- - -) vs. $\mathbb{N}(\mu_i(z_1), \sigma_i^2(z_1))$ (-·-·) vs. Beta$(\mu_i(z_1), \sigma_i(z_1))$ (——).

Outline

Estimation & Inference

- Full-information marginal maximum likelihood estimation $\Longrightarrow \hat{\Theta} = \arg\max \ell(\Theta; \mathbf{y})$.
- **Computation**: Sequential implementation:
  - EM-algorithm (fixed number of iterations/convergence):
    - ✓ Monotonic increments, easy implementation.
    - ✗ Slow: (sub-)linear convergence rate.
  - Direct maximisation via (quasi-)Newton algorithm (refinement step):
    - ✓ Fast: (super-)linear convergence rate.
    - ✗ Computationally intensive.
  - Numerical integration via Gauss-Hermite quadrature
- **Inference**: $\sqrt{n}(\hat{\Theta} - \Theta^*) \xrightarrow{d} \mathbb{N}(\mathbf{0}, (\mathcal{I}/n)^{-1})$, with expected information matrix $\mathcal{I}$.

Outline

Example 1: PISA 2018

- Computer-based Maths exam, booklet 1 (Brazil, $n = 1280$).
- Study of the 'Speed-accuracy trade-off' (SAT) (van der Linden, 2007; Molenaar et al., 2015).

- Confirmatory joint model for item response (y_i's) and response times (t_i's).
- Latent ability (z_1) and Latent 'speed' trait (z_2), correlated: $\text{Corr}(z_1, z_2) \neq 0$.
- Item responses (x9), $y_i \,|\, \mathbf{z} \sim \text{Bernoulli}(\pi_i(z_1))$
- Response times (x9, log-minutes):
  - Option 1 (current literature): $\log(t_i) \,|\, \mathbf{z} \sim \mathbb{N}(\mu_i(z_2), \sigma_i^2(z_2))$.
  - Option 2: $\log(t_i) \,|\, \mathbf{z} \sim \text{SN}(\mu_i(z_2), \sigma_i^2(z_2), \nu_i(z_2))$.

Example 1: PISA 2018

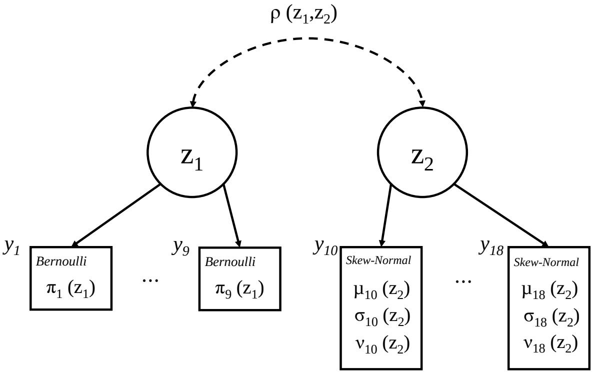


Figure: Path diagram example: Joint IR and RT model

Example 1: PISA 2018

	Model	AIC	BIC	K
1.	Bernoulli (π) + Normal (μ, fixed $\alpha_{i1,\mu}$)	26173.08	26368.96	38
2.	Bernoulli (π) + Normal (μ)	25908.67	26145.79	46
3.	Bernoulli (π) + Normal (μ, σ)	25754.91	26038.42	55
4.	Bernoulli (π) + Skew-Normal (μ)	25326.02	25609.53	55
5.	Bernoulli (π) + Skew-Normal (μ, σ)	25281.41	25611.30	64
6.	Bernoulli (π) + Skew-Normal (μ, ν)	25232.80	25562.70	64
7.	Bernoulli (π) + Skew-Normal (μ, σ, ν)	**25171.90**	**25548.18**	73

Table: Confirmatory GLVM-LSS for item responses and response times.

Example 1: PISA 2018

Item	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	(0.06)	0.79	(0.09)	0.19	(0.01)	-0.17	(0.01)	-0.95	(0.02)	-0.03	(0.02)	0.63	(0.13)	-0.03	(0.15)
Item 2	-0.47	(0.07)	1.03	(0.10)	0.30	(0.01)	-0.24	(0.01)	-0.89	(0.02)	-0.10	(0.02)	1.56	(0.15)	-0.78	(0.18)
Item 3	-0.04	(0.08)	1.95	(0.20)	0.42	(0.02)	-0.25	(0.01)	-0.61	(0.02)	-0.08	(0.02)	-1.07	(0.14)	0.81	(0.16)
Item 4	-0.69	(0.07)	0.96	(0.10)	0.45	(0.01)	-0.34	(0.01)	-0.87	(0.01)	-0.05	(0.02)	-1.45	(0.12)	-0.33	(0.15)
Item 5	-2.84	(0.21)	2.28	(0.24)	1.00	(0.02)	-0.36	(0.02)	-0.68	(0.02)	0.14	(0.02)	-1.04	(0.13)	-0.32	(0.20)
Item 6	-0.91	(0.06)	0.32	(0.08)	0.16	(0.01)	-0.36	(0.01)	-0.97	(0.02)	0.03	(0.03)	0.11	(0.11)	-0.88	(0.21)
Item 7	-4.79	(0.42)	2.49	(0.32)	0.65	(0.01)	-0.33	(0.01)	-1.15	(0.02)	-0.04	(0.02)	0.35	(0.14)	-0.58	(0.16)
Item 8	-3.67	(0.30)	2.39	(0.28)	1.02	(0.01)	-0.39	(0.01)	-1.02	(0.02)	0.12	(0.02)	-1.22	(0.17)	-1.60	(0.26)
Item 9	-2.73	(0.16)	1.46	(0.16)	0.58	(0.01)	-0.30	(0.01)	-0.90	(0.02)	-0.00	(0.02)	0.30	(0.10)	0.36	(0.13)

Est. Correlation(z_1, z_2) = -0.28 (SE: 0.025)

Table: Results for joint model of item responses and response times (Model 7).

Example 1: PISA 2018

Item	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	(0.06)	0.79	(0.09)	0.19	(0.01)	-0.17	(0.01)	-0.95	(0.02)	-0.03	(0.02)	0.63	(0.13)	-0.03	(0.15)
Item 2	-0.47	(0.07)	1.03	(0.10)	0.30	(0.01)	-0.24	(0.01)	-0.89	(0.02)	-0.10	(0.02)	1.56	(0.15)	-0.78	(0.18)
Item 3	-0.04	(0.08)	1.95	(0.20)	0.42	(0.02)	-0.25	(0.01)	-0.61	(0.02)	-0.08	(0.02)	-1.07	(0.14)	0.81	(0.16)
Item 4	-0.69	(0.07)	0.96	(0.10)	0.45	(0.01)	-0.34	(0.01)	-0.87	(0.01)	-0.05	(0.02)	-1.45	(0.12)	-0.33	(0.15)
Item 5	-2.84	(0.21)	2.28	(0.24)	1.00	(0.02)	-0.36	(0.02)	-0.68	(0.02)	0.14	(0.02)	-1.04	(0.13)	-0.32	(0.20)
Item 6	-0.91	(0.06)	0.32	(0.08)	0.16	(0.01)	-0.36	(0.01)	-0.97	(0.02)	0.03	(0.03)	0.11	(0.11)	-0.88	(0.21)
Item 7	-4.79	(0.42)	2.49	(0.32)	0.65	(0.01)	-0.33	(0.01)	-1.15	(0.02)	-0.04	(0.02)	0.35	(0.14)	-0.58	(0.16)
Item 8	-3.67	(0.30)	2.39	(0.28)	1.02	(0.01)	-0.39	(0.01)	-1.02	(0.02)	0.12	(0.02)	-1.22	(0.17)	-1.60	(0.26)
Item 9	-2.73	(0.16)	1.46	(0.16)	0.58	(0.01)	-0.30	(0.01)	-0.90	(0.02)	-0.00	(0.02)	0.30	(0.10)	0.36	(0.13)

Est. Correlation(z_1, z_2) = -0.28 (SE: 0.025)

Table: Results for joint model of item responses and response times (Model 7).

Example 1: PISA 2018

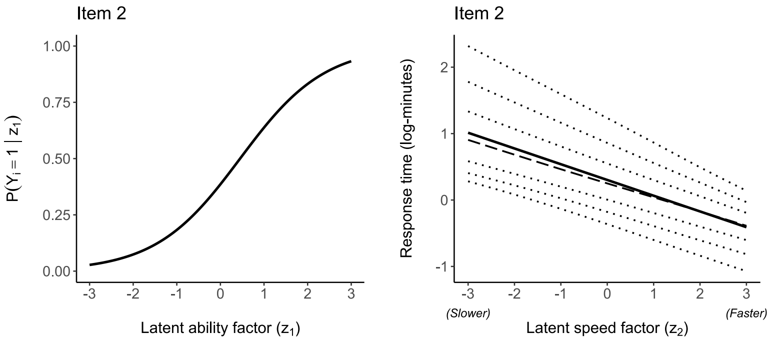


Figure: Item 2: Conditional expected values (——), median (– – –), and percentiles (······).

Example 1: PISA 2018

Item	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	(0.06)	0.79	(0.09)	0.19	(0.01)	-0.17	(0.01)	-0.95	(0.02)	-0.03	(0.02)	0.63	(0.13)	-0.03	(0.15)
Item 2	-0.47	(0.07)	1.03	(0.10)	0.30	(0.01)	-0.24	(0.01)	-0.89	(0.02)	-0.10	(0.02)	1.56	(0.15)	-0.78	(0.18)
Item 3	-0.04	(0.08)	1.95	(0.20)	0.42	(0.02)	-0.25	(0.01)	-0.61	(0.02)	-0.08	(0.02)	-1.07	(0.14)	0.81	(0.16)
Item 4	-0.69	(0.07)	0.96	(0.10)	0.45	(0.01)	-0.34	(0.01)	-0.87	(0.01)	-0.05	(0.02)	-1.45	(0.12)	-0.33	(0.15)
Item 5	-2.84	(0.21)	2.28	(0.24)	1.00	(0.02)	-0.36	(0.02)	-0.68	(0.02)	0.14	(0.02)	-1.04	(0.13)	-0.32	(0.20)
Item 6	-0.91	(0.06)	0.32	(0.08)	0.16	(0.01)	-0.36	(0.01)	-0.97	(0.02)	0.03	(0.03)	0.11	(0.11)	-0.88	(0.21)
Item 7	-4.79	(0.42)	2.49	(0.32)	0.65	(0.01)	-0.33	(0.01)	-1.15	(0.02)	-0.04	(0.02)	0.35	(0.14)	-0.58	(0.16)
Item 8	-3.67	(0.30)	2.39	(0.28)	1.02	(0.01)	-0.39	(0.01)	-1.02	(0.02)	0.12	(0.02)	-1.22	(0.17)	-1.60	(0.26)
Item 9	-2.73	(0.16)	1.46	(0.16)	0.58	(0.01)	-0.30	(0.01)	-0.90	(0.02)	-0.00	(0.02)	0.30	(0.10)	0.36	(0.13)

Est. Correlation(z_1, z_2) = -0.28 (SE: 0.025)

Table: Results for joint model of item responses and response times (Model 7).

Example 1: PISA 2018

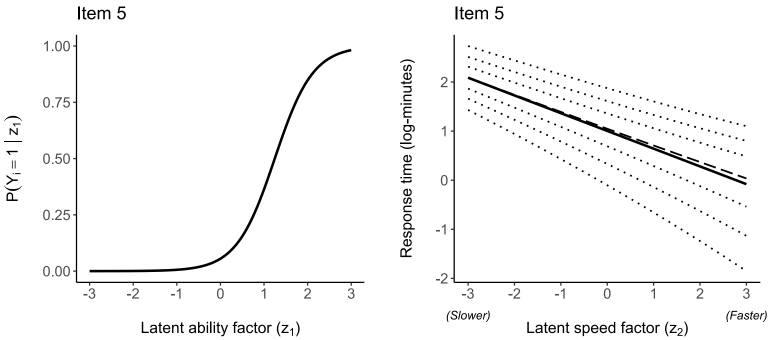


Figure: Item 5: Conditional expected values (——), median (- - -), and percentiles (······).

Example 2: ANES 2020

- American National Election Study (ANES): Thermometer questions (scaled ratings to $(0, 1)$ interval, $n = 7253$).
- Q: "*From 0 (cold) to 100 (hot), how would you rate* ______ *?*"

- Exploratory model on attitudes towards social groups and movements.
- Latent construct: 'progressive-conservative' scale (z_1).
- Items (x13): $y_i \mid z_1 \sim \text{Beta}(\mu_i(z_1), \sigma_i(z_1))$.

Example 2: ANES 2020

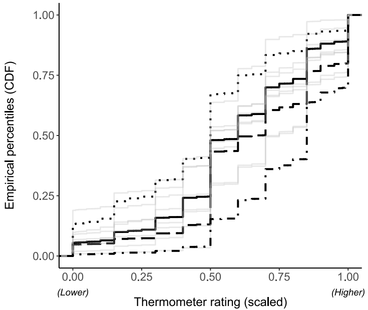


Figure: Empirical CDF: Feminists (——), Gay men and Lesbians (– – –), Christian fundamentalists (······), and Scientists (–·–·).

Model	AIC	BIC	K
Beta (μ)	-95075.12	-94806.44	39
Beta (μ, σ)	**-96805.52**	**-96447.28**	52

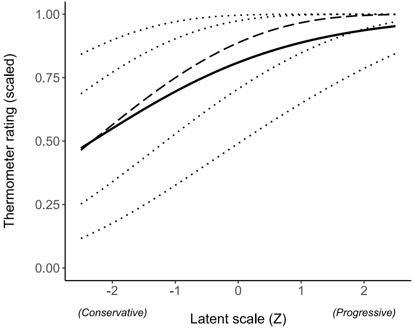
Table: Beta GLVM-LSS results for ANES 2020 dataset

Example 2: ANES 2020

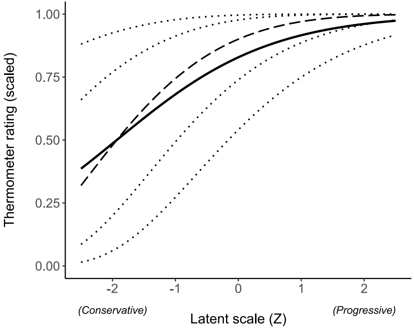
Item	Location parameter (μ)				Scale parameter (σ)			
	$\alpha_{i0,\mu}$		$\alpha_{i1,\mu}$		$\alpha_{i0,\sigma}$		$\alpha_{i1,\sigma}$	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Christian fundament.	-0.19	(0.02)	-0.47	(0.02)	0.67	(0.01)	0.05	(0.01)
Christians	0.96	(0.02)	-0.26	(0.02)	0.59	(0.01)	0.07	(0.01)
Muslims	0.41	(0.01)	0.98	(0.02)	0.01	(0.01)	-0.10	(0.01)
Jews	1.15	(0.02)	0.51	(0.02)	0.29	(0.01)	-0.16	(0.01)
Gay men and Lesbians	0.90	(0.02)	1.31	(0.02)	-0.06	(0.01)	-0.27	(0.01)
Transgender people	0.55	(0.01)	1.37	(0.02)	-0.12	(0.02)	-0.17	(0.01)
Feminists	0.45	(0.01)	1.21	(0.02)	-0.10	(0.01)	-0.16	(0.01)
#MeeToo movement	0.41	(0.02)	1.26	(0.02)	0.15	(0.02)	-0.28	(0.01)
BLM movement	0.21	(0.02)	1.23	(0.02)	0.53	(0.01)	-0.33	(0.01)
Labour Unions	0.39	(0.01)	0.62	(0.01)	0.21	(0.01)	-0.13	(0.01)
Big Businesses	-0.17	(0.01)	-0.14	(0.01)	0.26	(0.01)	0.05	(0.01)
Journalists	0.02	(0.01)	0.93	(0.02)	0.23	(0.01)	-0.17	(0.01)
Scientists	1.58	(0.02)	0.82	(0.02)	0.02	(0.01)	-0.25	(0.01)

Table: Results for the heteroscedastic Beta factor model.

Example 2: ANES 2020



(a) Homoscedastic model



(b) Heteroscedastic model

Figure: Item: Scientists, conditional expected values (——), median (- - -), and percentiles (······).

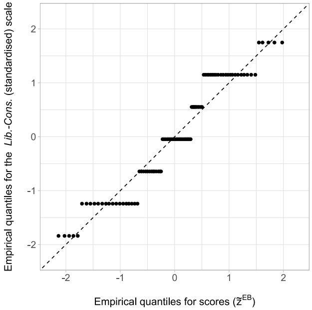
Example 2: ANES 2020

- Empirical Bayes (EB) factor scores:

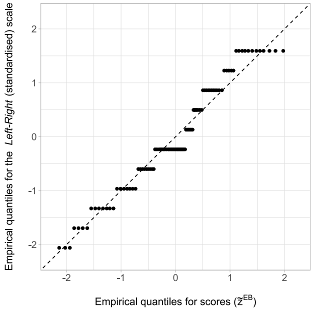
$$\tilde{\mathbf{z}}_m^{\mathsf{EB}} = \mathbb{E}(\mathbf{z} \,|\, \mathbf{y}_m; \hat{\Theta}) = \int_{\mathbb{R}^q} \mathbf{z} \cdot p(\mathbf{z} \,|\, \mathbf{y}_m; \hat{\Theta}) \; \mathsf{d}\mathbf{z} = \int_{\mathbb{R}^q} \mathbf{z} \cdot \frac{f(\mathbf{y}_m \,|\, \mathbf{z}; \hat{\Theta}_y) \; p(\mathbf{z}; \hat{\Theta}_z)}{\int_{\mathbb{R}^q} f(\mathbf{y}_m \,|\, \mathbf{z}'; \hat{\Theta}_y) \; p(\mathbf{z}'; \hat{\Theta}_z) \; \mathsf{d}\mathbf{z}'} \; \mathsf{d}\mathbf{z}$$

- Robustness check vs. self-reported measure of political orientation
  - 7-point (1 to 7) scale: Liberal vs. Conservative (LC)
  - 11-point (0 to 10) scale: Left vs. Right (LR)
- Correlations: Corr(EB,LC) = 0.65, Corr(EB,LR) = 0.56.

Example 2: ANES 2020



(a) *Liberal-Conservative* scale



(b) *Left-Right* scale

Figure: QQ-plots: (standardised) political orientation scales vs. EB factor scores (sign reversed).

Outline

Summary:

- We propose a GLVM for Location, Scale and Shape parameters (GLVM-LSS), that allows for modelling items with distributions beyond the exponential family and higher order moments as functions of the latent factors, under either exploratory and confirmatory settings.
- Model parameters are estimated using a two-step marginal maximum likelihood estimation procedure.
- We test GLVM-LSS and its estimation framework with some examples using survey data.
- Extensions: New distributions, (non-linear) additive measurement equations.
- **Future research**: local model fit criteria (residuals?), better ways for dealing with latent variables, penalised estimation for better interpretation and sparse solutions.

Thank you!

Camilo A. Cárdenas-Hurtado

@: c.a.cardenas-hurtado@lse.ac.uk

: @ccardehu

: https://github.com/ccardehu/GLVM-LSS

Appendix

Appendix: Details on EM-algorithm I

- Augmented/complete-data log-likelihood:

$$\ell_c(\Theta; \mathbf{y}, \mathbf{z}) = \sum_{m=1}^{n} \log f(\mathbf{y}_m, \mathbf{z}_m; \Theta) = \sum_{m=1}^{n} \left[\left\{ \sum_{i=1}^{p} \log f_i(y_{im} \,|\, \mathbf{z}; \boldsymbol{\theta}_i) \right\} + \log p(\mathbf{z}_m; \boldsymbol{\Phi}) \right]$$

- **E-step**: $\mathcal{Q}(\Theta; \Theta^{[t]}) = \mathbb{E}_{\mathbf{z} \,|\, \mathbf{y}; \Theta^{[t]}} \left[\ell_c(\Theta; \mathbf{y}, \mathbf{z}) \right]$
- **M-step**: $\Theta^{[t+1]} = \arg\max \mathcal{Q}(\Theta; \Theta^{[t]})$, or it suffices that $\mathcal{Q}(\Theta^{[t+1]}; \Theta^{[t]}) \geq \mathcal{Q}(\Theta^{[t]}; \Theta^{[t]})$.
  - Computation: $\mathbb{S}^{[t]} := \nabla_{\Theta} \mathcal{Q}(\Theta; \Theta^{[t]}) = \mathbf{0}$
  - NR update rule: $\Theta^{[t+1]} = \Theta^{[t]} - (\mathbb{H}^{[t]})^{-1} \, \mathbb{S}^{[t]}$, with $\mathbb{H}^{[t]} := \nabla_{\Theta} \nabla_{\Theta^\intercal} \mathcal{Q}(\Theta; \Theta^{[t]})$.

Appendix: Details on EM-algorithm II

- Score vector (entries):

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

- Observed information matrix (block diagonal matrix with entries):

$$\mathbb{H}^{[t]}_{[\bar{k}_{i,\varphi}, \bar{k}_{i,\tilde{\varphi}}]} = \sum_{m=1}^{n} \int_{\mathbb{R}^q} \left[\frac{\partial^2 \log f_i(y_{im} \,|\, \mathbf{z})}{\partial \boldsymbol{\alpha}_{i,\varphi} \partial \boldsymbol{\alpha}_{i,\tilde{\varphi}}^{\mathsf{T}}} \right] p(\mathbf{z} \,|\, \mathbf{y}_m; \Theta^{[t]}) \; \mathsf{d}\mathbf{z}$$

Appendix: Details on direct MLE via (quasi-)Newton algorithm

- Score vectors are equivalent: $\nabla_{\Theta}\ell(\Theta;\mathbf{y}) \equiv \nabla_{\Theta}\mathcal{Q}(\Theta;\Theta^{[t]}) = \mathbb{S}^{[t]}$ (Louis, 1982).
- For trust-region algorithm, $\mathcal{H}^{[t]} = \nabla_{\Theta}\nabla_{\Theta^{\intercal}}\ell(\Theta;\mathbf{y})$:

$$
\begin{aligned}
\mathcal{H}^{[t]}_{[\bar{k}_{i,\varphi},\bar{k}_{i',\tilde{\varphi}}]} &= \sum_{m=1}^{n}\int_{\mathbb{R}^q} p(\mathbf{z}\,|\,\mathbf{y}_m)\cdot\frac{\partial^2 \log f_i(y_{im}\,|\,\mathbf{z})}{\partial\boldsymbol{\alpha}_{i,\varphi}\partial\boldsymbol{\alpha}^{\intercal}_{i',\tilde{\varphi}}}\,\mathrm{d}\mathbf{z} \\
&+ \sum_{m=1}^{n}\int_{\mathbb{R}^q} p(\mathbf{z}\,|\,\mathbf{y}_m)\cdot\frac{\partial \log f_i(y_{im}\,|\,\mathbf{z})}{\partial\boldsymbol{\alpha}_{i,\varphi}}\cdot\frac{\partial \log f_{i'}(y_{i'm}\,|\,\mathbf{z})}{\partial\boldsymbol{\alpha}^{\intercal}_{i',\tilde{\varphi}}}\,\mathrm{d}\mathbf{z} \\
&- \sum_{m=1}^{n}\int_{\mathbb{R}^q} p(\mathbf{z}\,|\,\mathbf{y}_m)\cdot\frac{\partial \log f_i(y_{im}\,|\,\mathbf{z})}{\partial\boldsymbol{\alpha}_{i,\varphi}}\,\mathrm{d}\mathbf{z}\cdot\int_{\mathbb{R}^q} p(\mathbf{z}\,|\,\mathbf{y}_m)\frac{\partial \log f_{i'}(y_{i'm}\,|\,\mathbf{z})}{\partial\boldsymbol{\alpha}^{\intercal}_{i',\tilde{\varphi}}}\,\mathrm{d}\mathbf{z}
\end{aligned}
$$

Appendix: Details on computation of factor correlations

- $\mathbf{\Phi}$ is a correlation matrix $\Longrightarrow$ positive semi-definiteness and diagonal entries equal to one.
- Cholesky decomposition: $\mathbf{\Phi} = \mathrm{L}\mathrm{L}^\intercal$, with j rows L_j.
- Solve $\nabla_{\mathrm{L}_j} \mathcal{Q}(\Theta; \Theta^{[t]}) \equiv \nabla_{\mathrm{L}_j} \ell(\Theta; \mathbf{y}) = \mathbb{S}^{[t]}_{\mathrm{L}_j} = \mathbf{0}$

$$\mathbb{S}^{[t]}_{\mathrm{L}_{j,[k]}} = -n \cdot \mathrm{tr}\left(\mathrm{L}^\intercal (\mathrm{L}\mathrm{L}^\intercal)^{-1} \mathrm{D}_{jk}\right) + \sum_{m=1}^{n} \left[\mathrm{tr}\left(\mathrm{G}_{jk} \mathbb{V}^{[t]}_m\right) + \breve{\mathbf{z}}^{[t]\intercal}_m \mathrm{G}_{jk} \breve{\mathbf{z}}^{[t]}_m\right]$$

- Projection step:

$$\mathrm{L}^{[t+1]}_j = \underset{\mathrm{L}_j : ||\mathrm{L}_j|| = 1}{\arg\min} ||\mathrm{L}_j - \tilde{\mathrm{L}}^{[t+1]}_j|| = \frac{1}{||\tilde{\mathrm{L}}^{[t+1]}_j||} \tilde{\mathrm{L}}^{[t+1]}_j, \qquad \text{for } j = 1, ..., q$$

Appendix: Model Selection

- A GLVM-LSS model: $\mathcal{M} = \{\Theta,\ \mathbf{z},\ \mathcal{F}\}$
  - Model parameters: $\Theta \in \Xi$
  - Latent variables: $\mathbf{z} \in \mathbb{R}^q$
  - Parametric distributions: $\mathcal{F} = \{f_1(\cdot \mid \mathbf{z}; \boldsymbol{\theta}_1), ..., f_p(\cdot \mid \mathbf{z}; \boldsymbol{\theta}_p)\}$
- **Nested models**: $\mathcal{M}_0 = \{\Theta \in \Xi_0, \mathbf{z}, \mathcal{F}\}$ and $\mathcal{M}_1 = \{\Theta \in \Xi_1, \mathbf{z}, \mathcal{F}\}$
  - LRT (Normal data), χ^2-test (binary/categorical data)

    ⚠ Sensitive to departures from distributional assumptions
- **Non-nested models**: $\mathcal{M}_0 = \{\Theta \in \Xi_0, \mathbf{z} \in \mathbb{R}^{q_0}, \mathcal{F}_0\}$ and $\mathcal{M}_1 = \{\Theta \in \Xi_1, \mathbf{z} \in \mathbb{R}^{q_1}, \mathcal{F}_1\}$,
  - Information Criteria (AIC, BIC)

References I

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.

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.

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.

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

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.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability (37). Boca Ratón, FL, US: Chapman & Hall / CRC.

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.

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.

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

References II

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.

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.

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

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.