

Exploratory and Confirmatory Factor Analysis Identification by Spherical Geometry

Haziq Jamil

Håvard Rue

A nice abstract goes here.

1 Motivation

Why identification choices matter? Consider the Gaussian factor model

$$\begin{aligned}y &= \Lambda\eta + \epsilon \\ \eta &\sim \mathcal{N}_q(0, \Psi) \\ \epsilon &\sim \mathcal{N}_p(0, \Theta)\end{aligned}$$

with $\eta \propto \epsilon$. The implied covariance matrix is

$$\Sigma_y := \text{Var}(y) = \Lambda\Psi\Lambda^\top + \Theta.$$

In Bayesian SEM (and especially approximate Bayesian SEM: Laplace, skew-normal marginals, two-piece Gaussians, etc.), the local geometry of the posterior around its mode matters enormously because the approximation quality is driven by the Hessian (curvature) at or near the mode. Although this work has implications for maximum likelihood (affects standard errors) and sampling based Bayesian inference (affects easiness of sampling from posterior state space).

The **standard SEM advice** is

1. Marker/anchoring: set one loading per factor (e.g., $\lambda_{i^*} = 1$); or
2. Standardise latent variables, i.e. fix latent variance $\psi = 1$ (per factor), estimating all loadings.

Textbooks often say these are “equivalent up to rescaling”. That statement is true at the level of fit (same Σ_y), but it hides a critical nuance:

Even if two parameterisations are theoretically equivalent, they can induce very different local curvature (Hessians) in the coordinates used by an optimiser / Laplace approximation. The mode may map across parameterisations, but the shape of the posterior in those coordinates can change dramatically.

This is not just philosophical. Empirically saw cases where Laplace-based approximations were “near-perfect” under one parameterisation (`std.lv = TRUE`) and noticeably worse under another (anchored + nonlinear covariance transforms), despite targeting the same underlying model.

2 Literature Review

In recent years, some papers have been published that become aware of this issue: that selecting anchors or scale identification should not be arbitrary. Here is a selection of relevant papers:

[1] K. A. Bollen, A. G. Lilly, and L. Luo. "Selecting Scaling Indicators in Structural Equation Models (Sems)." In: *_Psychological Methods_* 29.5 (Oct. 2024), pp. 868–889. ISSN: 1939-1463, 1082-989X. DOI: 10.1037/met0000530.

[2] B. Graves and E. C. Merkle. "A Note on Identification Constraints and Information Criteria in Bayesian Latent Variable Models". In: *_Behavior Research Methods_* 54.2 (Aug. 2021), pp. 795–804. ISSN: 1554-3528. DOI: 10.3758/s13428-021-01649-8.

[3] E. Klopp and S. Klößner. "Scaling Metric Measurement Invariance Models". In: *_Methodology_* 19.3 (Sep. 2023), pp. 192–227. ISSN: 1614-2241. DOI: 10.5964/meth.10177.

[4] E. Klopp and S. Klößner. "The Impact of Scaling Methods on the Properties and Interpretation of Parameter Estimates in Structural Equation Models with Latent Variables". In: *_Structural Equation Modeling: A Multidisciplinary Journal_* 28.2 (Mar. 2021), pp. 182–206. ISSN: 1070-5511, 1532-8007. DOI: 10.1080/10705511.2020.1796673.

[5] T. D. Little, D. W. Slegers, and N. A. Card. "A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models". In: *_Structural Equation Modeling: A Multidisciplinary Journal_* 13.1 (Jan. 2006), pp. 59–72. ISSN: 1070-5511, 1532-8007. DOI: 10.1207/s15328007sem1301_3.

[6] J. H. Steiger. "When Constraints Interact: A Caution about Reference Variables, Identification Constraints, and Scale Dependencies in Structural Equation Modeling." In: *_Psychological Methods_* 7.2 (2002), pp. 210–227. ISSN: 1939-1463, 1082-989X. DOI: 10.1037/1082-989X.7.2.210.

3 Case: 1-factor model

In the 1-factor case, a scaling symmetry and ridge geometry emerges. Write $\Psi = \psi$ (a scalar), $\Lambda \in \mathbb{R}^{p \times 1}$. Then

$$\Sigma_y = \psi \Lambda \Lambda^\top + \Theta.$$

There is an exact scaling invariance: for any $c > 0$,

$$(\Lambda, \psi) \mapsto (c\Lambda, \psi/c^2) \Rightarrow \psi \Lambda \Lambda^\top = (\psi/c^2) (c\Lambda)(c\Lambda)^\top.$$

So the likelihood (and posterior absent informative priors that break this symmetry) is flat along a radial ridge in the joint (Λ, ψ) directions: only the product $\psi \Lambda \Lambda^\top$ is identified, not the individual “scale split” between ψ and Λ .

This ridge is easy to see in profile likelihood plots: high density follows approximately

$$\psi \propto \frac{1}{\lambda_i^2} \iff \log \psi = \text{const} - 2 \log |\lambda_i|.$$

Different identification constraints are different “cuts” through this ridge, and those cuts can be more or less orthogonal to the ridge depending on which loading is reliable/noisy.

Key intuition

Think of a long mountain ridge (likelihood high along a curve). Fixing $\lambda_{i^*} = 1$ is a vertical slice; fixing $\psi = 1$ is a horizontal slice. Because the ridge is curved and its steepness depends on item reliability, some slices pass through a sharp peak (good curvature), others pass through a broad flat region (ill conditioning).

3.1 Anchoring

Anchoring can be geometrically unstable because it chooses an indicator and forces it to be “special”. If that indicator is weak/noisy, the posterior can be highly anisotropic: the ridge becomes shallow in the anchored coordinate system, producing an ill-conditioned Hessian.

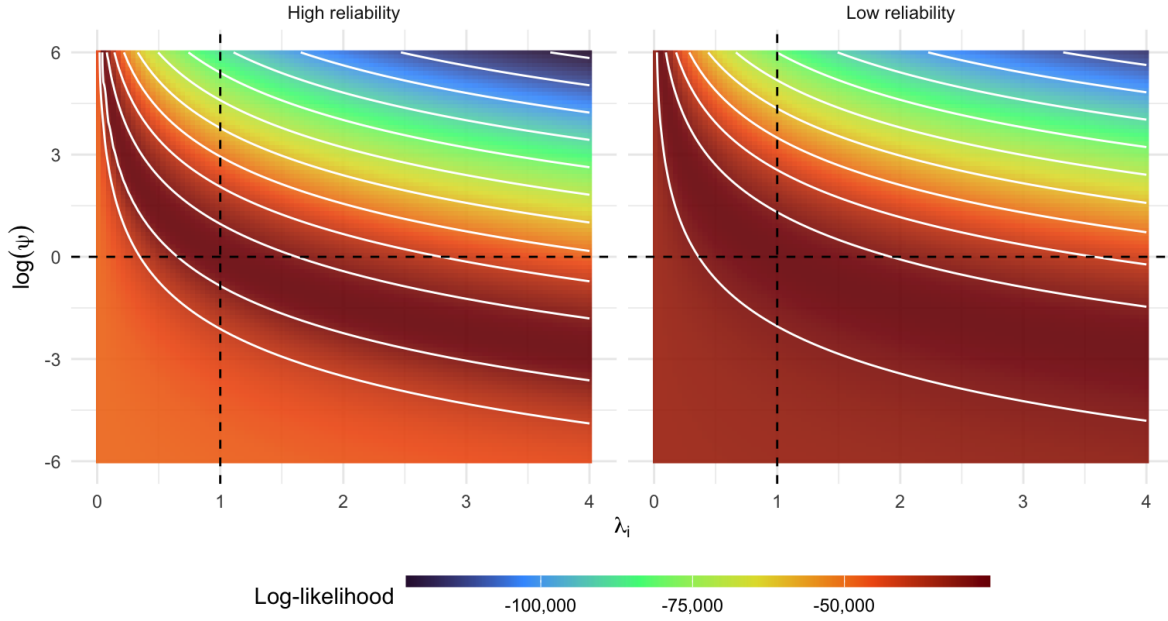
This leads to:

- poor curvature estimates (large condition numbers),

- degraded Laplace approximations (since Laplace is curvature-driven),
- skewed or heavy-tailed marginal shapes that simple Gaussian/Skew-Normal fits approximate less well.

We also observed empirically (via repeated simulations and condition numbers):

- anchoring on strong indicators often yields good conditioning,
- anchoring on weak indicators can be catastrophic,
- you typically do not know “the gold standard item” a priori, so anchoring is risky as a default.



3.2 A spherical reparameterisation

A principled way to respect the scaling symmetry is to explicitly separate:

- a direction (pattern of loadings), and
- a radial scale (overall factor strength).

Write

$$\Lambda = r \tilde{\Lambda}, \quad r > 0, \quad \|\tilde{\Lambda}\| = 1,$$

and treat r (or r^2) as the scale parameter. Then

$$\psi \Lambda \Lambda^\top = \psi r^2 \tilde{\Lambda} \tilde{\Lambda}^\top.$$

We can absorb scale either into ψ or into r . A clean choice is to let the latent variance carry the scale and fix $\|\Lambda\| = 1$ (equivalently set $r = 1$):

$$\|\Lambda\| = 1, \quad \psi \text{ free} \quad \Rightarrow \quad \Sigma_y = \psi \Lambda \Lambda^\top + \Theta.$$

Now:

- Λ is a direction on the unit sphere \mathcal{S}^{p-1} ,
- ψ is the radial strength (how much common variance the factor contributes).

This removes the scale ridge by construction.

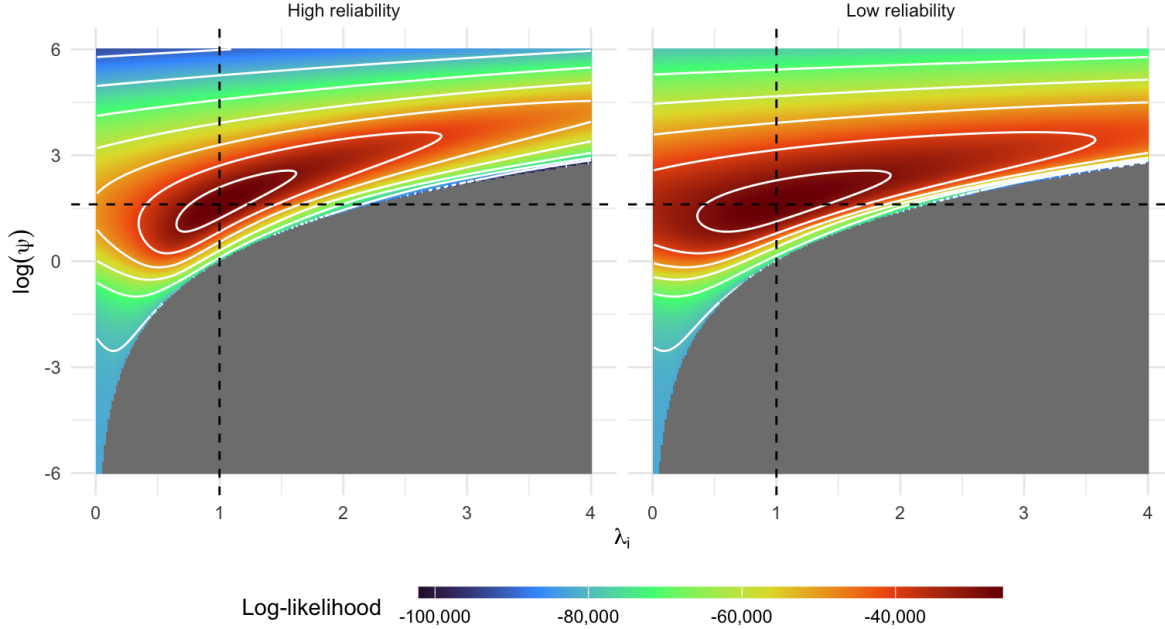
A natural prior emerges if we place a spherically symmetric prior on Λ , e.g. $\Lambda \sim \mathcal{N}(0, \tau^2 I)$. Writing $\Lambda = ru$ with $u \in \mathcal{S}^{p-1}$ yields a density that factorises into a radial part and a directional part. The key consequence:

- the induced prior on the direction u is uniform on the sphere (no preferred orientation),
- all directional structure must come from the likelihood (or from an explicit directional prior if desired).

This is one of the conceptual appeals: it aligns with “prior ignorance about orientation” in a way that anchoring does not.

Remark

Elias remarked that we can also naturally place PC priors in this framework, by defining the base model as the no-factor model ($r = 0$) which yields an exponential distribution on r .



3.3 Planar (effects-coding)

We also studied an alternative that is not spherical: effects coding (a “planar” constraint). For each factor column λ_j ,

$$\mathbf{1}^\top \lambda_j = c_j$$

often $c_j = p_j$ or $c_j = 0$ after centring. Geometrically, this constrains λ_j to an affine hyperplane (flat manifold) orthogonal to $\mathbf{1}$. This:

- avoids choosing a special marker item,
- breaks the scale symmetry in a linear way,
- tends to produce very well-behaved curvature (excellent condition numbers) in independent-cluster CFA.

Empirically, in repeated simulations for independent-cluster CFA, we found:

- effects coding (“planar”) often had the best conditioning,
- spherical was typically second best,
- fixing $\psi = 1$ was often okay,
- anchoring was highly sensitive to whether the marker item was strong/weak.

Interpretation: planar is a flat, symmetric “coordinate cut” through the ridge, whereas spherical is a curved symmetry-respecting cut. Both can improve geometry; planar can be surprisingly strong for independent clusters.

4 Multi-factor CFA

In multifactor CFA, there are two types of Λ matrix designs.

1. Independent-cluster: each item loads on only one factor (simple structure).
2. Complex structure: items can load on multiple factors, including the most complex scenario where all items load on all factors (Λ is dense) as in EFA.

4.1 Independent-cluster CFA

Let Λ be block-sparse by design (each item loads on only one factor). Two practical geometric approaches:

4.1.1 A. Product-of-spheres (per factor block)

For factor j with loading block $\lambda_j \in \mathbb{R}^{p_j}$,

$$\|\lambda_j\| = 1 \quad \text{for each } j,$$

and let Ψ be free (e.g., parameterise $\Psi = D^{1/2}CD^{1/2}$ with log-variances and an LKJ prior on C). This is a product manifold

$$\mathcal{S}^{p_1-1} \times \dots \times \mathcal{S}^{p_m-1},$$

which is simple and works well because clusters are structurally separated.

4.1.2 B. Planar / effects coding per factor block

For each factor block,

$$\mathbf{1}^\top \lambda_j = c_j$$

(plus a sign convention). This is a masked affine subspace, easy to optimise in Euclidean coordinates, and in our experiments it often produced the best conditioning.

Takeaway for independent-cluster CFA

You can get excellent geometry without “true” Stiefel machinery because the CFA sparsity already breaks much of the problematic interaction.

4.2 Complex structure CFA

When cross-loadings appear (or in EFA where Λ is dense), there is rotational indeterminacy:

For any orthogonal $R \in \mathbb{R}^{m \times m}$ with $R^\top R = I$,

$$\tilde{\Lambda} = \Lambda R, \quad \tilde{\Psi} = R^\top \Psi R \quad \Rightarrow \quad \tilde{\Lambda} \tilde{\Psi} \tilde{\Lambda}^\top = \Lambda \Psi \Lambda^\top.$$

So the likelihood is flat along a rotation ridge as well.

Appeal to a Stiefel parameterisation. A common geometric move is to represent the loading “direction frame” as

$$Q \in \mathcal{V}_{p,m} = \{Q : Q^\top Q = I\},$$

and write

$$\Sigma_y = Q \Psi Q^\top + \Theta,$$

with $\Psi \succ 0$ parameterised by a Cholesky factor (and optionally decomposed into scales + correlations for LKJ priors).

This is “rotation-respecting”: it does not force an arbitrary coordinate frame like effects coding does. Instead, it treats the space of orthonormal frames as the natural domain for “directions”.

In exploratory factor analysis (EFA), the likelihood depends on the loading matrix Λ only through the covariance component $\Lambda \Psi \Lambda^\top$. Consequently, for any orthogonal rotation $R \in O(m)$,

$$(\Lambda, \Psi) \mapsto (\Lambda R, R^\top \Psi R) \implies \Lambda \Psi \Lambda^\top = (\Lambda R)(R^\top \Psi R)(\Lambda R)^\top,$$

so the model-implied covariance Σ_y is unchanged. This is the familiar rotational indeterminacy: (Λ, Ψ) is not identified as a unique pair.

A useful distinction is between:

- Resolving (fixing) rotations: adding a convention (e.g., varimax, oblimin, or sign/order rules) that selects a single representative from the equivalence class $\{(\Lambda R, R^\top \Psi R) : R \in O(m)\}$. This yields interpretability, but it is fundamentally a choice of gauge rather than information supplied by the likelihood.
- Respecting rotations: parameterising the model so the rotational symmetry is expressed explicitly, without privileging an arbitrary coordinate frame during optimisation.

The Stiefel/Grassmann perspective formalises the latter. Write the “directional” part of the loadings as $Q \in \mathbb{R}^{p \times m}$ with orthonormal columns,

$$Q^\top Q = I_m \quad (Q \in \mathcal{V}_{p,m}, \text{ the Stiefel manifold}),$$

and let all scale and dependence be carried by a free positive definite matrix $\Psi \succ 0$. Rotations act as

$$(Q, \Psi) \mapsto (QR, R^\top \Psi R),$$

which leaves $\Sigma_y = Q\Psi Q^\top + \Theta$ invariant. Importantly, this action stays within the Stiefel manifold: if $Q \in \mathcal{V}_{p,m}$ then $QR \in \mathcal{V}_{p,m}$. Thus, Stiefel coordinates do not “fix” rotational indeterminacy; they respect it by representing the symmetry in a geometrically natural way.

From an identifiability standpoint, what EFA identifies is not Q itself but the m -dimensional subspace spanned by its columns:

$$\text{span}(Q) \in \text{Gr}(m, p),$$

a point on the Grassmann manifold $\text{Gr}(m, p)$. The mapping $Q \mapsto QR$ changes the basis but not the subspace, so all matrices in the orbit $\{QR : R \in O(m)\}$ correspond to the same point on $\text{Gr}(m, p)$. Varimax and related rotations can then be viewed as a post hoc gauge-fixing step that chooses an interpretable basis within the same subspace—without altering the fitted covariance Σ_y .

This viewpoint clarifies a common practice in EFA: “fix latent variances and rotate”. Fixing latent variances (e.g., $\Psi = I$) is primarily a scaling convention, while varimax is a rotation convention. The Stiefel formulation separates these roles cleanly: it removes scaling ambiguity by construction (scale is carried by Ψ , while Q is direction-only), and it treats rotations as an explicit symmetry to be handled by a transparent gauge choice (e.g., varimax) rather than implicitly by the initial parametrisation.

However, we notice that effects coding is not appropriate for EFA. Effects coding imposes column-wise linear constraints (hyperplanes). Those constraints are not rotation-invariant: if you rotate columns, the constraints generally fail. Hence, effects coding can be excellent for CFA (where rotation is not a symmetry because the mask fixes orientation), but it is not a principled solution for EFA’s inherent rotational symmetry.

4.3 Proposal to handle sparsity when using Stiefel-style estimation

Manifold optimisation (dense Q) tends to lose sparsity automatically. So while it helps EFA, other design matrices in the context of CFA would not benefit. We considered two ways to impose CFA structure:

4.3.1 A. Hard zeros via a masked submanifold (more complex)

Restrict Q to satisfy both:

$$Q^\top Q = I, \quad Q_{ij} = 0 \text{ if mask } M_{ij} = 0.$$

This is a “masked Stiefel” constraint. It enforces zeros during estimation, but projections/retractions become more delicate.

4.3.2 B. Dense Stiefel estimation + post-hoc rotation to a hypothesised mask

Estimate dense Q and Ψ stably, then find an orthogonal rotation R that best matches the hypothesised CFA structure M , e.g. via a Procrustes-type objective:

$$R^* = \arg \min_{R^\top R = I} \|(QR) \odot (1 - M)\|_F^2,$$

and transform:

$$\tilde{Q} = QR^*, \quad \tilde{\Psi} = R^{*\top} \Psi R^*.$$

This preserves Σ_y exactly (hence fit), while producing loadings in a CFA-interpretable orientation. Importantly:

- off-mask nonzeros measure mismatch between your hypothesised structure and what the data support,
- exact zeros are not guaranteed unless you hard-threshold or refit with mask constraints.

5 Discussion

Why “it doesn’t matter” is incomplete

It’s true that many parameterisations are equivalent at the level of Σ_y , but for Laplace-type methods:

- the approximation quality depends on curvature in the chosen coordinates,
- anchoring (or other ad hoc cuts) can create ill-conditioned Hessians in realistic regimes (weak markers, certain correlation structures),
- “recoverability” of a target parameterisation by algebraic rescaling does not guarantee the approximate posterior (or standard errors) is equally accurate across parameterisations.

Practical recommendation (CFA focus)

- Independent clusters (no cross-loadings): use planar (effects coding) or per-factor spherical normalisation for stable geometry; avoid anchoring unless you truly have a gold-standard marker.
- Cross-loadings: use dense Stiefel-style geometry for estimation stability, then rotate post-hoc to a hypothesised mask (or use masked Stiefel if exact zeros are essential).
- EFA: use Stiefel + rotation (varimax/oblimin etc.). Effects coding is not principled for EFA because it breaks rotation symmetry.

Novelty

- Not novel: “standardised loadings” effects coding, and rotation methods exist.

- Novel angle we developed: treating identification as a geometry problem for approximate Bayesian inference, showing that:
 1. identification choices reshape curvature and can materially affect Laplace quality;
 2. symmetry-respecting reparameterisations (spherical/Stiefel) can stabilise optimisation and approximation;
 3. effects coding can be viewed as optimisation on a flat affine submanifold (“planar geometry”), explaining its empirical conditioning benefits in independent-cluster CFA;
 4. for cross-loadings, a practical workflow is “good geometry first (dense Stiefel), interpretability second (mask-aligned rotation).”

- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed). Wiley.
- Bollen, K. A. (1989). *Structural equations with latent variables* (pp. xiv, 514). John Wiley & Sons. <https://doi.org/10.1002/9781118619179>
- Bollen, K. A., Lilly, A. G., & Luo, L. (2024). Selecting scaling indicators in structural equation models (sems). *Psychological Methods*, 29(5), 868–889. <https://doi.org/10.1037/met0000530>
- Graves, B., & Merkle, E. C. (2021). A note on identification constraints and information criteria in Bayesian latent variable models. *Behavior Research Methods*, 54(2), 795–804. <https://doi.org/10.3758/s13428-021-01649-8>
- Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs*, 48, xi + 91–xi + 91.
- Jöreskog, K. G., & Moustaki, I. (2001). Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivariate Behavioral Research*, 36(3), 347–387. <https://doi.org/10.1207/S15327906347-387>
- Klopp, E., & Klößner, S. (2021). The Impact of Scaling Methods on the Properties and Interpretation of Parameter Estimates in Structural Equation Models with Latent Variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(2), 182–206. <https://doi.org/10.1080/10705511.2020.1796673>
- Klopp, E., & Klößner, S. (2023). Scaling metric measurement invariance models. *Methodology*, 19(3), 192–227. <https://doi.org/10.5964/meth.10177>
- Klößner, S., & Klopp, E. (2019). Explaining Constraint Interaction: How to Interpret Estimated Model Parameters Under Alternative Scaling Methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1), 143–155. <https://doi.org/10.1080/10705511.2018.1517356>
- Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Wiley.
- Little, T. D., Slegers, D. W., & Card, N. A. (2006). A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3
- Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2), 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>

- Song, X., & Lee, S. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences* (1st ed.). Wiley. <https://doi.org/10.1002/9781118358887>
- Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, 7(2), 210–227. <https://doi.org/10.1037/1082-989X.7.2.210>