

# Limited information goodness-of-fit tests for ordinal factor models

Haziq Jamil

Limited information approaches overcome sparsity issues and computational challenges in traditional goodness-of-fit tests. This paper describes the implementation of LIGOF tests for ordinal factor models that have been fitted using the `{lavaan}` package in R. The tests are computationally efficient and reliable, and adapted to suit whichever parameter estimation procedure was used to fit the model. The implementation is available as an R package called `{lavaan.ligof}`.

## 1 Introduction

- Focus on limited information methods that use up to second-order moments of the data.
- This synergises well with the LIGOF tests which also use up to second-order moments.
- **IF** full information tests are used, then there is still the computational burden of computing the full multinomial matrix  $\Sigma$  which grows exponentially with the number of variables. Using limited information methods to estimate the parameters offers a way to avoid this.
- Besides, most software uses limited information methods to estimate the parameters of ordinal factor models, such as the `{lavaan}` package in R, Mplus, Stata, and LISREL.
- Sometimes, GLS methods involve a fit function which are themselves asymptotically chi square, and this can be used for testing fit. However, more popular versions use thresholds and polychoric correlations, and in this case it is not possible to detect sources of misfit.
- It would appear that calculation of LIGOF tests statistics involve quantities that are already computed in the process of estimating the parameters of the model, so it is not computationally burdensome to compute these tests.

Table 1: Response patterns for  $p = 3$  with  $m_1 = 2$ , and  $m_2 = m_3 = 3$ .

| $r$ | $y_1$ | $y_2$ | $y_3$ | Pattern |
|-----|-------|-------|-------|---------|
| 1   | 1     | 1     | 1     | 111     |
| 2   | 1     | 1     | 2     | 112     |
| 3   | 1     | 1     | 3     | 113     |
| 4   | 1     | 2     | 1     | 121     |
| 5   | 1     | 2     | 2     | 122     |
| 6   | 1     | 2     | 3     | 123     |
| 7   | 1     | 3     | 1     | 131     |
| 8   | 1     | 3     | 2     | 132     |
| 9   | 1     | 3     | 3     | 133     |

  

| $r$ | $y_1$ | $y_2$ | $y_3$ | Pattern |
|-----|-------|-------|-------|---------|
| 10  | 2     | 1     | 1     | 211     |
| 11  | 2     | 1     | 2     | 212     |
| 12  | 2     | 1     | 3     | 213     |
| 13  | 2     | 2     | 1     | 221     |
| 14  | 2     | 2     | 2     | 222     |
| 15  | 2     | 2     | 3     | 223     |
| 16  | 2     | 3     | 1     | 231     |
| 17  | 2     | 3     | 2     | 232     |
| 18  | 2     | 3     | 3     | 233     |

Source: [Article Notebook](#)

Source: [Article Notebook](#)

## 2 Methods

### 2.1 Ordinal data

Consider the case of analysing multivariate data  $\mathbf{y} = (y_1, \dots, y_p)^\top$ , where each item  $y_i$  is an ordinal random variable with  $m_i$  categories,  $i = 1, \dots, p$ . Let  $\mathcal{R} = \{\mathbf{c} = (c_1, \dots, c_p)^\top \mid c_i \in \{1, \dots, m_i\}\}$  be the set of all possible response patterns, and let  $R = \prod_i m_i$  be the cardinality of this set. The joint probability of observing a response pattern  $\mathbf{c}_r \in \mathcal{R}$  is given by

$$\pi_r = \Pr(\mathbf{y} = \mathbf{c}_r) = \Pr(y_1 = \mathbf{c}_{r1}, \dots, y_p = \mathbf{c}_{rp}), \quad r = 1, \dots, R, \quad (1)$$

with  $\sum_r \pi_r = 1$ . Collect all response probabilities into the vector  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_R)^\top \in [0, 1]^R$ . An example with  $p = 3$ ,  $m_1 = 2$ , and  $m_2 = m_3 = 3$  is given below. In total, there are  $R = 2 \times 3 \times 3 = 18$  response patterns as shown in Table 1.

Later on we wish to use lower-order residuals to assess the fit of a model to the data, which first requires a description of lower-order moments and its connection to the joint response probabilities. Marginally, each  $y_i$  can be viewed as a multinoulli random variable with event probabilities  $\pi_k^{(i)} = \Pr(y_i = k)$ ,  $k = 1, \dots, m_i$ , that sum to one. Therefore, this univariate distribution is characterised by its  $(m_i - 1)$  *moments*  $\pi_2^{(i)}, \dots, \pi_{m_i}^{(i)}$ , with the first moment being redundant due to the sum to unity constraint. All univariate moments can be collected into the vector  $\dot{\boldsymbol{\pi}}_1 = (\pi_k^{(i)})^\top$  whose dimension is  $S_1 = \sum_i (m_i - 1)$ . In a similar light, the bivariate distribution of  $(y_i, y_j)$  is characterised by its  $(m_i - 1)(m_j - 1)$  *joint moments*  $\pi_{k,l}^{(ij)} = \Pr(y_i = k, y_j = l)$ ,  $k = 2, \dots, m_i$ ,  $l = 2, \dots, m_j$ . Also collect all bivariate moments into the vector  $\dot{\boldsymbol{\pi}}_2 = (\pi_{k,l}^{(ij)})^\top$  whose dimension is  $S_2 = \sum_{i < j} (m_i - 1)(m_j - 1)$ . Finally, denote by  $\boldsymbol{\pi}_2 = (\dot{\boldsymbol{\pi}}_1^\top, \dot{\boldsymbol{\pi}}_2^\top)^\top$  the vector of multivariate moments up to order 2, which is a vector of length  $S = S_1 + S_2$ .

Because the lower order moments are contained in the higher order moments, the vector  $\pi_2$  can be extracted from the joint probabilities  $\pi$  via a linear operation  $\pi_2 = \mathbf{T}_2\pi$  (Jamil et al., 2025). As an example, continuing from the  $p = 3$  instance above, the moments for the first variable  $y_1$ ,  $\Pr(y_1 = 2)$  can be obtained by *summing* over all joint probabilities whose patterns contain  $y_1 = 2$ . The positions of these joint probabilities in the vector  $\pi$  are picked up by the first row of the matrix  $\mathbf{T}_2$ . Similarly, the two bivariate moments of  $(y_1, y_2)$ , i.e.  $\pi_{22}^{(12)}$  and  $\pi_{23}^{(12)}$  are obtained by summing over the joint probabilities whose patterns contain  $y_1 = 2$  and  $y_2 = 2$ , and  $y_1 = 2$  and  $y_2 = 3$ , respectively.

|            | 111 | 112 | 113 | 121 | 122 | 123 | 131 | 132 | 133 | 211 | 212 | 213 | 221 | 222 | 223 | 231 | 232 | 233 |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Y1=2       | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| Y2=2       | 0   | 0   | 0   | 1   | 1   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 0   | 0   | 0   |
| Y2=3       | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   |
| Y3=2       | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   |
| Y3=3       | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   |
| Y1=2, Y2=2 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 0   | 0   | 0   |
| Y1=2, Y2=3 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   |
| Y1=2, Y3=2 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   | 0   |
| Y1=2, Y3=3 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 1   | 0   | 0   | 1   |
| Y2=2, Y3=2 | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   |
| Y2=2, Y3=3 | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   |
| Y2=3, Y3=2 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   |
| Y2=3, Y3=3 | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   |

Source: [Article Notebook](#)

Figure 1: Matrix  $\mathbf{T}_2$  for the case of  $p = 3$  with  $m_1 = 2$ , and  $m_2 = m_3 = 3$ .

Note that this construction of lower-order moments generalises to any order  $q \leq p$ , but the total number of moments up to order  $q$  grows combinatorially in both  $p$  and the category counts  $m_i$ , yielding design matrices  $\mathbf{T}_q$  that can become computationally burdensome. Moreover, although we arbitrarily dropped the first moment in the foregoing construction, the choice of which category to omit is immaterial. This is because category probabilities sum to one, so excluding any one category produces a similar-dimensional parameterisation algebraically equivalent to excluding any other. For further details, consult Reiser (1996) and Maydeu-Olivares & Joe (2006).

## 2.2 Confirmatory factor analysis

The confirmatory factor analysis (CFA) model imposes a structure on the joint response probabilities by assuming that the  $p$  observed variables are manifestations of a smaller set of  $q$  latent variables. In this way, the CFA may be viewed as a data-reduction technique since,

effectively, the correlations among variables are modelled by a pre-specific factor structure using lower-dimensional data summaries.

CFA is typically used for continuous manifest variables, but it can also be applied to ordinal data. A common approach is the *underlying variable* (UV) approach, where the observed responses  $y_i$  are assumed to be discretised versions of continuous latent variables  $y_i^*$ . The connection is made through

$$y_i = \begin{cases} 1 & \tau_0^{(i)} < y_i^* < \tau_1^{(i)} \\ 2 & \tau_1^{(i)} < y_i^* < \tau_2^{(i)} \\ 3 & \tau_2^{(i)} < y_i^* < \tau_3^{(i)} \\ \vdots & \vdots \\ m_i & \tau_{m_i-1}^{(i)} < y_i^* < \tau_{m_i}^{(i)}, \end{cases}$$

with the *thresholds*  $\tau_k^{(i)}$  for item  $i$  satisfying the ordering

$$-\infty \equiv \tau_0^{(i)} < \tau_1^{(i)} < \tau_2^{(i)} < \dots < \tau_{m_i-1}^{(i)} < \tau_{m_i}^{(i)} \equiv +\infty.$$

Evidently, the model is invariant to a linear transformation, since scaling and shifting the underlying variables  $y_i^*$  do not affect the outcome of the ordinal variable  $y_i$ . For this reason it is convenient to assume, for the purposes of model identifiability, a zero mean Gaussian distribution  $\mathbf{y}^* \sim N_p(\mathbf{0}, \Sigma_{\mathbf{y}^*})$ , where  $\Sigma_{\mathbf{y}^*}$  is a correlation matrix.

The underlying continuous variables  $\mathbf{y}^*$ , unlike their discrete counterparts  $\mathbf{y}$ , are now suitable to be modelled using a factor analysis model. Here, the goal is to find a set of latent factors  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_q)^\top \in \mathbb{R}^q$ , with  $q \ll p$ , that sufficiently explain the covariance structure of the  $p$ -dimensional variable space. This is achieved by the relationship

$$\mathbf{y}^* = \mathbf{\Lambda} \boldsymbol{\eta} + \boldsymbol{\epsilon},$$

where  $\mathbf{\Lambda}$  is a (often sparse)  $p \times q$  matrix of factor loadings, and  $\boldsymbol{\epsilon}$  is a vector of residuals. Certain distributional assumptions are made, namely that  $\boldsymbol{\eta} \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$  with  $\boldsymbol{\Psi}$  a correlation matrix,  $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Theta}_\epsilon)$  with  $\boldsymbol{\Theta}_\epsilon = \mathbf{I} - \text{diag}(\mathbf{\Lambda} \boldsymbol{\Psi} \mathbf{\Lambda}^\top)$ , and that  $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\epsilon}) = \mathbf{0}$ . Together, this implies that the polychoric correlation matrix of  $\mathbf{y}$  is given by

$$\Sigma_{\mathbf{y}^*} = \mathbf{\Lambda} \boldsymbol{\Psi} \mathbf{\Lambda}^\top + \boldsymbol{\Theta}_\epsilon \in \mathbb{R}^{p \times p}.$$

As a remark, the UV approach is commonly employed in the context of confirmatory factor analysis (CFA) models due to the ease of modelling, though other approaches such as item response theory (IRT) models are also available (Jöreskog & Moustaki, 2001).

For this factor analysis model, the parameters of interest are the non-zero entries  $\boldsymbol{\lambda}$  of the loading matrix  $\mathbf{\Lambda}$ , the unique non-diagonal entries  $\boldsymbol{\psi}$  in the factor correlation matrix  $\boldsymbol{\Psi}$ , and the thresholds  $\boldsymbol{\tau}^{(i)} = (\tau_1^{(i)}, \dots, \tau_{m_i-1}^{(i)})^\top$  for each ordinal item  $y_i$ . Collectively, these parameters

are denoted by  $\theta = (\boldsymbol{\lambda}^\top, \boldsymbol{\rho}^\top, \boldsymbol{\tau}^{(1)}, \dots, \boldsymbol{\tau}^{(p)})^\top$  belonging to some parameter space  $\Theta$ . Under this CFA model, each joint response probability  $\pi_r$  from Equation 1 is now evaluated as a function of  $\theta$ :

$$\pi_r := \pi_r(\theta) = \int \cdots \int_{\mathcal{C}_r} \phi_p(\mathbf{y}^* \mid \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}^*}) \, d\mathbf{y}^*, \quad (2)$$

where the  $p$ -dimensional integral is taken over the set  $\mathcal{C}_r = \{\mathbf{y}^* \in \mathbb{R}^p \mid y_i = \mathbf{c}_{ri}, i = 1, \dots, p\}$ , the set of all continuous values that yield the response pattern  $\mathbf{c}_r$ .

## 2.3 Consistent estimators for $\theta$

Suppose that a sample  $\mathcal{Y} = \{\mathbf{y}^{(h)}\}_{h=1}^n$  is obtained, where  $\mathbf{y}^{(h)} = (y_1^{(h)}, \dots, y_p^{(h)})^\top$  represents the  $p$ -dimensional ordinal-data observation from subject  $h \in \{1, \dots, n\}$ . As a remark, samples may not necessarily be independent, and in such cases, corresponding sampling weights  $\omega_s$  can be used to account for the sampling design (Jamil et al., 2025), and most of what will be discussed below can be adapted to account for this.

Many methods exist to estimate the parameters  $\theta$  of the CFA model, but we are most interested in those that yield a  $\sqrt{n}$ -consistent and asymptotically normal estimator. Specifically, we assume that  $\hat{\theta}$  satisfies

$$\sqrt{n}(\hat{\theta} - \theta) = \hat{\mathbf{Q}} \cdot \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta)) + o_p(1), \quad (3)$$

where the term  $\mathbf{p} = (p_1, \dots, p_R)^\top$  is the vector of empirical joint response proportions, and  $\hat{\mathbf{Q}} \xrightarrow{P} \mathbf{Q}$  as  $n \rightarrow \infty$  is some *influence matrix* that performs asymptotic linearisation from the joint response proportions  $\mathbf{p}$  to the parameters  $\theta$ . This includes a wide range of likelihood-based (Bock & Lieberman, 1970; Lord, 1968) and pseudolikelihood-based (Alfonzetti et al., 2025; Katsikatsou et al., 2012) methods, as well as generalised least squares (GLS) based methods (Christofferson, 1975; Jöreskog, 1990, 1994; Jöreskog & Moustaki, 2001; Muthén, 1978, 1984), with GLS popularly implemented as a multi-stage estimation procedure in software. Equation 3 holds true whether full information methods (i.e., estimation using joint response probabilities) or limited information methods (i.e., using a lower-order subset of the response probabilities) are employed.

A neat way of viewing the parameter estimation is that most of these methods are a class of M-estimators. M-estimation provides a general and flexible framework for parameter estimation, in which estimators are obtained by minimizing an objective function  $F(\theta)$ , typically expressed as an empirical average  $\sum_{s=1}^n F(\mathbf{y}_s, \theta)$ , or, equivalently, by solving a system of estimating equations  $\sum_{s=1}^n \nabla_\theta F(\mathbf{y}_s, \theta) = \mathbf{0}$ , where  $\nabla_\theta F = \partial F / \partial \theta$ . This formulation encompasses a wide range of classical and robust procedures, including maximum likelihood, least squares, and weighted least squares methods mentioned above.

In the context of confirmatory factor analysis (CFA) with ordinal indicators, the estimating equations typically arise from a discrepancy function defined on thresholds and polychoric correlations, and M-estimation offers a principled way to derive estimators even when the full

likelihood is computationally intractable. A central assumption in this framework is that there exists a parameter  $\theta_0 \in \Theta$  such that the population moment condition  $E[\nabla_\theta F(\mathbf{y}, \theta_0)] = 0$  holds. This condition is not a consequence of the data, but rather a theoretical premise about the underlying data-generating mechanism. It defines the parameter value to which the estimator is expected to converge. In a correctly specified model,  $\theta_0$  corresponds to the true parameter; in the presence of misspecification, it instead represents the value that best satisfies the moment condition within the assumed model class.

Under standard regularity conditions—such as continuity of  $\nabla_\theta F$  in  $\theta$ , measurability, and uniform convergence of empirical averages—the M-estimator  $\hat{\theta}$  is consistent and asymptotically normal (Huber, 1964; van der Vaart, 1998). Specifically,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(\mathbf{0}, \mathbf{V}(\theta)),$$

where the asymptotic variance is given by the sandwich formula  $\mathbf{V}(\theta) = \mathcal{H}(\theta)^{-1} \mathcal{J}(\theta) \mathcal{H}(\theta)^{-T}$ , with

$$\mathcal{H}(\theta) = E \left[ -\nabla_\theta^2 F(\mathbf{y}, \theta) \right], \quad \mathcal{J}(\theta) = E \left[ \nabla_\theta F(\mathbf{y}, \theta) \nabla_\theta F(\mathbf{y}, \theta)^\top \right].$$

The matrix  $\mathcal{H}$  is known as the *sensitivity matrix* and is estimated consistently by  $\hat{\mathbf{H}} = -\frac{1}{n} \sum_{s=1}^n \nabla_\theta^2 F(\mathbf{y}_s, \hat{\theta})$ . The matrix  $\mathcal{J}$  is known as the *variability matrix* and is estimated consistently by  $\hat{\mathbf{J}} = \frac{1}{n} \sum_{s=1}^n \nabla_\theta F(\mathbf{y}_s, \hat{\theta}) \nabla_\theta F(\mathbf{y}_s, \hat{\theta})^\top$ .

These properties make M-estimation particularly appealing in settings where the data are ordinal and the working model may be misspecified, as is often the case in large-scale psychometric applications. For a detailed treatment of the asymptotic theory of M-estimators in econometric and semiparametric contexts, see Newey & McFadden (1994). For the commonly used techniques to estimate CFA, the table below gives an overview for the form that  $F$  and its derivatives take.

Source: [Article Notebook](#)

Achieving the desired form stated in Equation 3 requires an asymptotic linearisation argument. For CFA models, a general M-estimator  $\hat{\theta}$  for  $\theta$  is obtained by solving the set of estimating equations

$$U(\theta) := n \mathbf{D}(\theta)^\top \mathbf{W}_\theta (\mathbf{m} - \mu(\theta)) = 0$$

where  $\mathbf{m}$  is a vector of sample moments,  $\mu(\theta)$  is the vector of model-implied moments,  $\mathbf{D}(\theta)$  is the Jacobian of the model-implied moments with respect to  $\theta$ , and  $\mathbf{W}_\theta$  is a weight matrix which may or may not depend on the parameters. Table 2 summarises these quantities for different estimators most commonly used for CFA. Under correct model specification, the sensitivity matrix takes the form

$$\mathcal{H}(\theta) = \mathbf{D}(\theta)^\top \mathbf{W}_\theta \mathbf{D}(\theta).$$

A first-order Taylor expansion of  $U(\hat{\theta})$  around  $\theta$  with a little rearranging and multiplying through by  $\sqrt{n}$  gives

$$\sqrt{n}(\hat{\theta} - \theta) = \left[ -\frac{1}{n} \frac{\partial U(\theta)}{\partial \theta} \right]^{-1} \mathbf{D}(\theta)^\top \mathbf{W}_\theta \cdot \sqrt{n}(\mathbf{m} - \mu(\theta)) + o_p(1),$$

Table 2: Objective functions and their derivatives for different estimators.

|                   | $\mathbf{m} - \boldsymbol{\mu}(\theta)$                             | $\mathbf{W}^{-1}$   | $\mathbf{D}(\theta)$  |
|-------------------|---|---|---|
| ML <sup>1</sup>   | $\mathbf{p} - \boldsymbol{\pi}(\theta)$                             | $\text{diag}(\boldsymbol{\pi}(\theta))$   | $\partial \boldsymbol{\pi}(\theta) / \partial \theta$               |
| PML <sup>2</sup>  | $\mathbf{p}_{\text{pair}} - \boldsymbol{\pi}_{\text{pair}}(\theta)$ | $\text{diag}(\boldsymbol{\pi}_{\text{pair}}(\theta))$                           | $\partial \boldsymbol{\pi}_{\text{pair}}(\theta) / \partial \theta$ |
| UBN <sup>3</sup>  | $\mathbf{p}_2 - \boldsymbol{\pi}_2(\theta)$                         | $\text{diag}(\boldsymbol{\pi}_2(\theta))$                                       | $\partial \boldsymbol{\pi}_2(\theta) / \partial \theta$             |
| MCS <sup>4</sup>  | $\mathbf{p} - \boldsymbol{\pi}(\theta)$                             | $\boldsymbol{\Sigma}$   | $\partial \boldsymbol{\pi}(\theta) / \partial \theta$               |
| MCS2 <sup>4</sup> | $\mathbf{p}_2 - \boldsymbol{\pi}_2(\theta)$                         | $\boldsymbol{\Sigma}_2$   | $\partial \boldsymbol{\pi}_2(\theta) / \partial \theta$             |
| ULS <sup>5</sup>  | $\mathbf{s} - \boldsymbol{\sigma}(\theta)$                          | $\mathbf{I}$  | $\partial \boldsymbol{\sigma}(\theta) / \partial \theta$            |
| WLS <sup>5</sup>  | $\mathbf{s} - \boldsymbol{\sigma}(\theta)$                          | $\boldsymbol{\Gamma} = \text{limvar}(\mathbf{s} - \boldsymbol{\sigma}(\theta))$ | $\partial \boldsymbol{\sigma}(\theta) / \partial \theta$            |
| DWLS <sup>5</sup> | $\mathbf{s} - \boldsymbol{\sigma}(\theta)$                          | $\text{diag}(\boldsymbol{\Gamma})$  | $\partial \boldsymbol{\sigma}(\theta) / \partial \theta$            |

<sup>1</sup>Maximum likelihood (Lord 1968; Bock and Lieberman 1970)

<sup>2</sup>Pairwise maximum likelihood (Katsikatsou, Moustaki, Yang-Wallentin, and Jöreskog 2012).  $\mathbf{p}_{\text{pair}}$  and  $\boldsymbol{\pi}_{\text{pair}}(\theta)$  vectors of sample and model-implied pairwise response probabilities, respectively.

<sup>3</sup>Underlying bivariate normal approach (Jöreskog and Moustaki 2001)

<sup>4</sup>Minimum chi square (Christoffersson 1975; Muthén 1978)

<sup>5</sup>Unweighted, weighted, or diagonally weighted least squares (Muthén 1984; Jöreskog 1990, 1994).  $\mathbf{s}$  and  $\boldsymbol{\sigma}(\theta)$  vectors of sample and model-implied thresholds and polychoric correlations, respectively.

where the observed Hessian  $-\frac{1}{n}\partial U(\theta)/\partial\theta \xrightarrow{P} \mathcal{H}(\theta)$  as  $n \rightarrow \infty$ . Taking limits, we see the influence matrix for CFA shaping up to involve

$$\left[-\frac{1}{n}\frac{\partial U(\theta)}{\partial\theta}\right]^{-1} \mathbf{D}(\theta)^\top \mathbf{W}_\theta \xrightarrow{P} \mathcal{H}(\theta)^{-1} \mathbf{D}(\theta)^\top \mathbf{W}_\theta =: \tilde{\mathbf{Q}} \quad \text{as } n \rightarrow \infty.$$

For certain full-information estimators like ML and MCS,  $\tilde{\mathbf{Q}}$  fits in to be premultiplied to the the  $R$ -vector of moment differences  $\mathbf{m} - \boldsymbol{\mu}(\theta) = \mathbf{p} - \boldsymbol{\pi}(\theta)$ , and thus Equation 3 is satisfied. Incidentally, in the case of ML, the influence matrix is  $\mathbf{Q} = \mathcal{I}^{-1} \boldsymbol{\Delta} \mathbf{W} \in \mathbb{R}^{t \times R}$ , where  $\mathcal{I} = \boldsymbol{\Delta}^\top \mathbf{W}^{-1} \boldsymbol{\Delta}^\top$  is the unit Fisher information,  $\boldsymbol{\Delta} = (\partial \boldsymbol{\pi} / \partial \theta)$  is the Jacobian of the joint response probabilities with respect to the parameters, and  $\mathbf{W} = \text{diag}(\boldsymbol{\pi})$  is a diagonal matrix of the joint response probabilities, agreeing with results in Maydeu-Olivares & Joe (2005). It can be further shown that  $\mathbf{Q}$  simplifies to  $\boldsymbol{\Delta}^\top \mathcal{I}^{-1} \boldsymbol{\Delta}$ .

In other cases, we need to post multiply the influence matrix  $\tilde{\mathbf{Q}}$  by an appropriate matrix so that it is able to conform to a matrix-vector multiplication with the joint probabilities as per Equation 3. This depends on the vector of moment differences. Consider a transformation  $g : \mathbf{p} \mapsto \mathbf{m}$  that maps the joint response probabilities  $\mathbf{p}$  to the moments  $\mathbf{m}$  (and likewise for the model implied moments), and let  $\mathbf{G} := \partial g / \partial \mathbf{p}$  be the Jacobian of the transformation. When dealing with PML, UBN or MCS2, then the transformation is linear since the lower order moments are linear functions of the joint response probabilities. On the other hand, ULS, WLS, and DWLS methods specify a transformation that is not linear, where the joint probabilities are mapped to the thresholds and polychoric correlations. Such a transformation was described by Muthén (1978) in the context of dichotomous data, but extends to the case of ordinal data too. In any case,

$$\begin{aligned} \sqrt{n}(\mathbf{m} - \boldsymbol{\mu}(\theta)) &= \sqrt{n}(g(\mathbf{p}) - g(\boldsymbol{\pi}(\theta))) \\ &= \mathbf{G} \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta)) + o_p(1). \end{aligned} \tag{4}$$

Plugging this into the above equation lets us see the form of the influence matrix as  $\mathbf{Q} = \mathcal{H}(\theta)^{-1} \mathbf{D}(\theta)^\top \mathbf{W}_\theta \mathbf{G}$ .

When using limited information methods, it would be sufficient to consider the lower-order moments transformation  $g_2 : \mathbf{p}_2 \mapsto \mathbf{m}$  instead. For PML, UBN, and MCS2 this is clearly obvious. For ULS, WLS, and DWLS, this also makes sense because the thresholds and polychoric correlations are functions of univariate and bivariate moments respectively. Letting  $\mathbf{G}_2 := \partial g_2 / \partial \mathbf{p}_2$ , we have

$$\begin{aligned} \sqrt{n}(\mathbf{m} - \boldsymbol{\mu}(\theta)) &= \sqrt{n}(g(\mathbf{p}_2) - g(\boldsymbol{\pi}_2(\theta))) \\ &= \mathbf{G}_2 \sqrt{n}(\mathbf{p}_2 - \boldsymbol{\pi}_2(\theta)) + o_p(1) \\ &= \mathbf{G}_2 \mathbf{T}_2 \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta)) + o_p(1), \end{aligned} \tag{5}$$

and thus Equation 3 holds with the influence matrix  $\mathbf{Q} = \mathbf{Q}_2 \mathbf{T}_2$ , where  $\mathbf{Q}_2 = \mathcal{H}(\theta)^{-1} \mathbf{D}(\theta)^\top \mathbf{W}_\theta \mathbf{G}_2$ . Consequently, when using limited information methods to estimate the parameters of a CFA model, it is sufficient to consider only consistency in relation to univariate and bivariate probabilities. We will see that this is useful when we come to the topic of residuals.



## 2.4 Distribution of residuals

Let  $p_r = n_r/n$  be the  $r$ th entry of the  $R$ -vector of sample proportions  $\mathbf{p}$ , where  $n_r$  is the number of times the response pattern  $\mathbf{c}_r$  was observed in the sample  $\mathcal{Y}$ . The random vector  $\mathbf{n} = (n_1, \dots, n_R)^\top$  follows a multinomial distribution with parameters  $n$ ,  $R$ , and  $\boldsymbol{\pi}$ , with  $E(\mathbf{n}) = n\boldsymbol{\pi}$  and variance

$$\text{Var}(\mathbf{n}) = n(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^\top) = n\boldsymbol{\Sigma}.$$

It is widely known (Agresti, 2002) for iid samples that

$$\sqrt{n}(\mathbf{p} - \boldsymbol{\pi}) \xrightarrow{D} N_R(\mathbf{0}, \boldsymbol{\Sigma}) \quad (6)$$

as  $n \rightarrow \infty$ , which is a consequence of the central limit theorem. Note that this also works for the case of weighted samples in complex sampling designs, but  $\boldsymbol{\Sigma}$  need not take a multinomial form in such cases (Fuller, 2009).

Consider testing the composite null hypothesis of  $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\theta_0)$  against the alternative  $H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}(\theta_0)$ . To do so, use the univariate and bivariate residuals  $\hat{\mathbf{e}}_2 = \mathbf{T}_2(\mathbf{p} - \boldsymbol{\pi}(\hat{\theta})) = \mathbf{T}_2\hat{\mathbf{e}}$  as the basis for the test statistic. Now we derive the asymptotic distribution of this quantity. Write

$$\begin{aligned} \sqrt{n}\hat{\mathbf{e}} &= \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta_0)) - \sqrt{n}(\boldsymbol{\pi}(\hat{\theta}) - \boldsymbol{\pi}(\theta_0)) \\ &= \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta_0)) - \sqrt{n}\boldsymbol{\Delta}(\hat{\theta} - \theta_0) + o_p(1), \end{aligned}$$

where we had considered a Taylor expansion of  $\boldsymbol{\pi}(\hat{\theta})$  around  $\theta_0$  to get to the second line, and defined  $\boldsymbol{\Delta} = (\partial\boldsymbol{\pi}(\theta)/\partial\theta)$ . Now, for  $\sqrt{n}$ -consistent estimators satisfying Equation 3, we have that

$$\begin{aligned} \sqrt{n}\hat{\mathbf{e}} &= \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta_0)) - \boldsymbol{\Delta}\hat{\mathbf{Q}} \cdot \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta_0)) + o_p(1) \\ &= (\mathbf{I} - \boldsymbol{\Delta}\hat{\mathbf{Q}}) \cdot \sqrt{n}(\mathbf{p} - \boldsymbol{\pi}(\theta_0)) + o_p(1), \end{aligned}$$

so it is clear that  $\hat{\mathbf{e}}$  is asymptotically normal by the CLT (Equation 6). Let  $\lim\text{var}(\hat{\mathbf{e}}) = \boldsymbol{\Omega}$ . Then, since  $\hat{\mathbf{e}}_2 = \mathbf{T}_2\hat{\mathbf{e}}$ , the lower-order residuals are also asymptotically normal with zero mean and variance  $\boldsymbol{\Omega}_2 = \mathbf{T}_2\boldsymbol{\Omega}\mathbf{T}_2^\top$ . The full form of the asymptotic variance is given by

$$\boldsymbol{\Omega}_2 = \boldsymbol{\Sigma}_2 - \boldsymbol{\Delta}_2\mathbf{Q}\boldsymbol{\Sigma}\mathbf{T}_2^\top - \mathbf{T}_2\boldsymbol{\Sigma}\mathbf{Q}^\top\boldsymbol{\Delta}_2^\top + \boldsymbol{\Delta}_2\mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^\top\boldsymbol{\Delta}_2^\top, \quad (7)$$

where  $\boldsymbol{\Delta}_2 = \mathbf{T}_2\boldsymbol{\Delta}$ . See Maydeu-Olivares & Joe (2008) for further details, including the use of residuals from moments of up to order  $q < p$ .

In practice, when limited information introduces inconsistency between the estimation method and the quantities derived from it, potentially leading to misleading inferences or misinterpretation of model fit. methods are used to estimate the parameters, the estimation of  $\boldsymbol{\Omega}_2$  involves plugging in full information quantities such as fitted probabilities and Jacobians. This is less than ideal, since it introduces inconsistency between the estimation method and the quantities derived from it, potentially leading to misleading inferences or misinterpretation of model fit. Furthermore, quantities such as the multinomial covariance matrix  $\boldsymbol{\Sigma}$  becomes exponentially large in dimension as  $p$  increases, making it difficult to work with.

One solution is to consider the weaker  $\sqrt{n}$ -consistent condition for limited information estimators suggested by Equation 5, in which the influence matrix  $\mathbf{Q}_2$  is utilised. Since  $\mathbf{Q} = \mathbf{Q}_2 \mathbf{T}_2$ , Equation 7 will simplify to

$$\begin{aligned}\boldsymbol{\Omega}_2 &= \boldsymbol{\Sigma}_2 - \boldsymbol{\Delta}_2 \mathbf{Q}_2 \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_2 \mathbf{Q}_2^\top \boldsymbol{\Delta}_2^\top + \boldsymbol{\Delta}_2 \mathbf{Q}_2 \boldsymbol{\Sigma} \mathbf{Q}_2^\top \boldsymbol{\Delta}_2^\top \\ &= (\mathbf{I} - \boldsymbol{\Delta}_2 \mathbf{Q}_2) \boldsymbol{\Sigma}_2 (\mathbf{I} - \boldsymbol{\Delta}_2 \mathbf{Q}_2)^\top.\end{aligned}\tag{8}$$

where  $\boldsymbol{\Sigma}_2 = \mathbf{T}_2 \boldsymbol{\Sigma} \mathbf{T}_2^\top$  is the covariance matrix of the lower-order moments. Computationally this is more efficient as it uses only quantities involving uni and bivariate moments, which are much smaller in size than the full joint response probabilities.

## 2.5 Wald-type tests

Given as  $\hat{\mathbf{e}}_2 \xrightarrow{D} N_S(\mathbf{0}, \boldsymbol{\Omega}_2)$ , we can construct a Wald test statistic for the null hypothesis  $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\theta_0)$  as

$$X^2 = n \hat{\mathbf{e}}_2^\top \hat{\boldsymbol{\Omega}}_2^{-1} \hat{\mathbf{e}}_2,$$

where  $\hat{\boldsymbol{\Omega}}_2$  is a consistent estimator of  $\boldsymbol{\Omega}_2$ . This test statistic is asymptotically distributed as chi square under the null hypothesis, with degrees of freedom equal to  $S - t$ , i.e. the number of lower-order moments used in the test minus the number of parameters estimated.

The computational challenges here are in the estimation of  $\hat{\boldsymbol{\Omega}}_2$  as well as the inversion of the matrix. Addressing the second issue first, suppose an estimator  $\hat{\boldsymbol{\Omega}}_2$  is available, then the Moore-Penrose pseudoinverse  $\hat{\boldsymbol{\Omega}}_2^+$  can be computed using the singular value decomposition (SVD) of  $\hat{\boldsymbol{\Omega}}_2$ . This sidesteps any numerical instabilities that may occur when inverting the matrix directly, since the rank of  $\boldsymbol{\Omega}_2$  may be deficient (Reiser, 1996), although inversion can still be computationally challenging when the dimension  $S$  is large.

Jamil et al. (2025) instead proposed a diagonal Wald test, in which  $\text{diag}(\hat{\boldsymbol{\Omega}}_2)^{-1}$  is used instead of the full matrix inverse. Since inverting a diagonal matrix is straightforward compared to the full (pseudo) inverse, this is indeed computationally efficient. However, simulation studies show that this is not as powerful as the full Wald test, in the context of pairwise likelihood estimation of binary CFA models.

On the estimation of  $\boldsymbol{\Omega}_2$ , which involves estimation of the  $\mathbf{Q}$  matrix, which may be involved depending on the estimation method used. A very attractive proposal by Maydeu-Olivares and colleagues (Maydeu-Olivares & Joe, 2005, 2006, 2008) is to consider using a matrix  $\boldsymbol{\Xi}$  such that  $\boldsymbol{\Omega}_2$  is a generalised inverse of  $\boldsymbol{\Xi}$ , i.e.  $\boldsymbol{\Xi} = \boldsymbol{\Xi} \boldsymbol{\Omega}_2 \boldsymbol{\Xi}$ . By denoting  $\boldsymbol{\Delta}_{2,\pi}^\perp$  to be an  $S \times (S - t)$  orthogonal complement to  $\boldsymbol{\Delta}_{2,\pi}$  satisfying  $\boldsymbol{\Delta}_{2,\pi}^\perp \boldsymbol{\Delta}_{2,\pi}^\top = \mathbf{0}$ , it can be shown that  $X^2 = \hat{\mathbf{e}}_2^\top \hat{\boldsymbol{\Xi}} \hat{\mathbf{e}}_2$  converges to the Wald test statistic with similar degrees of freedom (Jamil et al., 2025), where

$$\boldsymbol{\Xi} = \boldsymbol{\Delta}_{2,\pi}^\perp ((\boldsymbol{\Delta}_{2,\pi}^\perp)^\top \boldsymbol{\Sigma}_2 \boldsymbol{\Delta}_{2,\pi}^\perp)^{-1} (\boldsymbol{\Delta}_{2,\pi})^\top.$$

This is advantageous in that it does not require the estimation of  $\mathbf{Q}$ , and only requires the Jacobian  $\Delta_{2,\pi}$  as well as a consistent estimator for  $\Sigma_2$ , which can be obtained from a plug-in estimator using the model-implied probabilities  $\pi(\hat{\theta})$ .

## 2.6 Pearson and general LIGOF tests

Wald-type tests may behave unstably and has poor small-sample behaviour (Jamil et al., 2025). As an alternative, a Pearson-type test can be constructed using the Pearson residuals

$$\begin{aligned} X^2 &= n \hat{\mathbf{e}}_2^\top \text{diag}(\pi_2(\hat{\theta}))^{-1} \hat{\mathbf{e}}_2 \\ &= n \sum_{i,k} \frac{p_k^{(i)} - \pi_k^{(i)}(\hat{\theta})}{\pi_k^{(i)}(\hat{\theta})} + n \sum_{i < j} \sum_{k < l} \frac{p_{k,l}^{(ij)} - \pi_{k,l}^{(ij)}(\hat{\theta})}{\pi_{k,l}^{(ij)}(\hat{\theta})}, \end{aligned}$$

where  $p_k^{(i)}$  and  $p_{k,l}^{(ij)}$  are the sample estimates for the univariate and bivariate response probabilities defined earlier. Similar test statistics were studied by Cai et al. (2006) and Bartholomew & Leung (2002), where the latter considered only bivariate margins. The Pearson test statistic does not follow an asymptotic chi-square distribution because of the dependence of the summands in the above equation. It does however converge to a sum of scaled chi-square variables  $\sum_{s=1}^S \delta_s Z_s$ , where each  $Z_s \stackrel{\text{iid}}{\sim} \chi_1^2$  and  $\delta_s$  are the eigenvalues of  $M = \Omega_2^{-1/2} \text{diag}(\pi_2(\hat{\theta}))^{-1} \Omega_2^{-1/2}$ .

For calculation of p-values, a moment matching procedure can be employed (Jamil et al., 2025; Maydeu-Olivares & Joe, 2008), where the first three moments of  $X^2$  are matched to the first three moments of some chi-square random variate, which is then used as the reference distribution to conduct the test. The moments of  $X^2$  are estimated using trace product formulae involving  $\text{diag}(\pi_2(\hat{\theta}))$  as well as  $\hat{\Omega}_2$ . Though the Pearson test looks as if the  $\Omega_2$  matrix is not required, it is actually required to compute the p-values.

More generally, any LIGOF test statistic can be constructed using  $X^2 = \hat{\mathbf{e}}_2^\top \hat{\Xi} \hat{\mathbf{e}}_2$ , where  $\hat{\Xi} \xrightarrow{D} \Xi$  is some  $S \times S$  weight matrix that can be arbitrarily chosen. We saw earlier that the Wald test involves  $\hat{\Xi} = \hat{\Omega}_2^+$ , while the Pearson test involves  $\hat{\Xi} = \text{diag}(\pi_2(\hat{\theta}))^{-1}$ . Other choices for this weight matrix are  $\hat{\Xi} = \mathbf{I}$  (RSS test) or  $\hat{\Xi} = \hat{\Sigma}_2^{-1}$  (Multinomial test).

## 3 Usage

## References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed). Wiley-Interscience.
- Alfonzetti, G., Bellio, R., Chen, Y., & Moustaki, I. (2025). Pairwise stochastic approximation for confirmatory factor analysis of categorical data. *British Journal of Mathematical and Statistical Psychology*, 78(1), 22–43. <https://doi.org/10.1111/bmsp.12347>

- Bartholomew, D. J., & Leung, S. O. (2002). A goodness of fit test for sparse 2p contingency tables. *British Journal of Mathematical and Statistical Psychology*, 55(1), 1–15.
- Bock, R. D., & Lieberman, M. (1970). Fitting a Response Model for  $n$  Dichotomously Scored Items. *Psychometrika*, 35(2), 179–197. <https://doi.org/10.1007/bf02291262>
- Cai, Li., Maydeu-Olivares, A., Coffman, D. L., & Thissen, David. (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2 tables. *British Journal of Mathematical and Statistical Psychology*, 59(1), 173–194. <https://doi.org/10.1348/000711005X66419>
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40(1), 5–32. <https://doi.org/10.1007/BF02291477>
- Fuller, W. A. (2009). *Introduction to statistical time series*. John Wiley & Sons.
- Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. <https://doi.org/10.1214/aoms/1177703732>
- Jamil, H., Moustaki, I., & Skinner, C. (2025). Pairwise likelihood estimation and limited-information goodness-of-fit test statistics for binary factor analysis models under complex survey sampling. *British Journal of Mathematical and Statistical Psychology*, 78(1), 258–285. <https://doi.org/10.1111/bmsp.12358>
- Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, 24(4), 387–404. <https://doi.org/10.1007/BF00152012>
- Jöreskog, K. G. (1994). On the Estimation of Polychoric Correlations and their Asymptotic Covariance Matrix. *Psychometrika*, 59(3), 381–389. <https://doi.org/10.1007/BF02296131>
- 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>
- Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, 56(12), 4243–4258. <https://doi.org/10.1016/j.csda.2012.04.010>
- Lord, F. M. (1968). An Analysis of the Verbal Scholastic Aptitude Test Using Birnbaum’s Three-Parameter Logistic Model. *Educational and Psychological Measurement*, 28(4), 989–1020. <https://doi.org/10.1177/001316446802800401>
- Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2 n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100(471), 1009–1020.
- Maydeu-Olivares, A., & Joe, H. (2006). Limited Information Goodness-of-fit Testing in Multidimensional Contingency Tables. *Psychometrika*, 71(4), 713–732. <https://doi.org/10.1007/s11336-005-1295-9>
- Maydeu-Olivares, A., & Joe, H. (2008). An overview of limited information goodness-of-fit testing in multidimensional contingency tables. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 253–262). Universal Academy Press.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43(4), 551–560. <https://doi.org/10.1007/BF02293813>

- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. <https://doi.org/10.1007/BF02294210>
- Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). Elsevier. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4)
- Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, 61(3), 509–528. <https://doi.org/10.1007/BF02294552>
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511802256>

## Acknowledgements

I thank Rabi'ah Roslan for her diligent contributions as part of her undergraduate project and for the insightful discussions that helped shape this paper.