A methodological perspective on the analysis of clinical and personality questionnaires

Smits, Iris Anna Marije

Chapter 5

On the Importance of the Skewness Parameter in Modeling Latent Traits

Abstract

The linear factor model for continuous items assumes normally distributed item scores. Deviations from normality due to non-normality of the expected item scores can be modeled through a skew-normally distributed factor model or a quadratic factor model. We show that the quadratic factor model is equivalent to the skew-normal factor model up to the third-order moments, but that the converse is not generally true. We illustrate that observed data that follow any skew-normal factor model can be well approximated by the quadratic factor model. We explain that this connotes the importance of including a skewness parameter in the latent trait for the establishment of measurement invariance across populations.

This chapter has been submitted for publication as:

Smits, I. A. M., Timmerman, M. E., & Stegeman, A. On the importance of the skewness parameter in modelling latent traits.

5.1 Introduction

The linear factor model for continuous items assumes normally distributed item scores. Non-normality of the item scores occurs in empirical practice, giving a need for alternatives to the linear factor model.

For continuous items, the normality assumption has been relaxed in different variants. The least restrictive alternatives are asymptotic distribution free factor analysis (Mooijaart, 1985), semi-parametric estimation (Ma & Genton, 2010) and nonparametric maximum likelihood estimation (Skrondal & Rabe-Hesketh, 2004, pp. 182-184). This type of analyses requires very large sample sizes, and is computationally heavy. Therefore, in empirical practice, parametric alternatives appear to be more attractive to account for non-normally distributed continuous items. The earliest approach is the nonlinear factor model (Mooijaart & Bentler, 1986), in which a normally distributed factor has a nonlinear relationship to the expected observed item scores. The nonlinearity is typically restricted to a quadratic function to make estimation feasible (e.g., Muthen & Muthen, 1998-2007; Molenaar, Dolan, & Verhelst, 2010). Recently, the skew-normal factor model (Molenaar et al., 2010; Montanari & Viroli, 2010) has been proposed, which pertains to a linear factor model with skewly distributed factors. Both the skew-normal factor model and the nonlinear factor model allow for non-normality of the expected item scores (i.e., apart from residuals). Non-normality of the residuals can be modeled as a function of the latent trait score, thereby allowing for heteroscedasticity of the residuals (Hessen & Dolan, 2009).

For ordinal polytomous items, non-normality can arise at the level of the latent continuous item scores underlying the observed item scores. That is, polytomous items are generally analyzed by means of an ordinal factor analysis (FA) model or the equivalent graded response model (GRM; Samejima, 1969; see Takane & de Leeuw, 1987). The ordinal FA model / GRM can be derived by assuming that the observed ordinal items relate to latent continuous items that are normally distributed, while the latent continuous item scores follow a linear factor model. The underlying normality assumptions can be relaxed by using the heteroscedastic GRM with a skewed latent trait (Molenaar, Dolan, & De Boeck, 2012) or, as a nonparametric alternative, the Ramsay-curve item response theory (IRT) model (e.g., Woods, 2006).

For continuous items, the parametric approaches to account for non-normality of the items scores have been put in a single framework (Molenaar et al., 2010). Of particular interest are the two variants that account for deviations from normality of the expected item scores. This is done either through a quadratic factor model, a nonlinear factor model with a polynomial of the second degree, or through a skew-normal factor model. On the basis of empirical identification checks (i.e., confidence intervals, simulation results and the rank of the Hessian), Molenaar et al. (2010) illustrated that the latter two variants cannot be implemented jointly into a single model. This implies that the quadratic factor model and the skew-normal factor model are competing in empirical practice. This raises the questions how they are related at a theoretical level, and how researchers should choose between them. When would either model be preferred, and when would the choice between the two models be arbitrary?

In this study, we consider and analyze the relation between the quadratic factor model and the skew-normal factor model as variants to account for deviations from normality of continuous expected item scores. Further, we discuss the implications of this relation for the choice between the variants, both in a single-group and in a multiple-group context. In particular, for a single-group, we first show that in the case of a skew-normal factor the two variants are empirically indistinguishable. We further show that the reverse does not generally hold, implying that some quadratic factor models could be empirically distinguished from a skew-factor model. We will argue that if both models would fit an empirical data set equally well, that the skew-normal factor model is preferred over the quadratic factor model, for reasons of model parsimony and ease of interpretation. Subsequently, we discuss the implications of those results in a multi-group context. In particular, we argue that in a multi-group context the skewness parameter is not a negligible parameter, but an important parameter to characterize a population g in comparison to g', with $g' \neq g$.

5.2 Modeling Non-Normal Expected Item Scores in a Single Population

5.2.1 The Skew-Normal Factor Model

The skew-normal factor model has been proposed as a valuable approach to account for non-normally distributed expected item scores. Molenaar and colleagues (2010) considered the skew-normal factor model for a single factor, while, independently, Montanari & Viroli (2010) considered the general model involving multiple factors.

We start with the skew-normal factor model for a single factor (Molenaar et al., 2010). If y_i denotes a randomly observed score on item *i*, the following linear factor model for y_i is specified:

$$y_i = \nu_i^* + \lambda_i^* \eta^* + \varepsilon_i^*, \tag{1}$$

where ν_i^* is the intercept of item *i*, λ_i^* is the factor loading of item *i*, η^* is the common factor score and ε_i^* is the residual. The residual is drawn from a normal distribution: $\varepsilon_i^* \sim N(0, \sigma^2)$, and the common factor score is drawn from a skew-normal distribution: $\eta^* \sim SN(\kappa, \omega, \zeta)$, with location parameter κ, scale parameter ω, and shape parameter ζ (as developed by Azzalini, 1985, 1986). The probability density function of the skew-normal factor scores η^* is the following:

$$f(\eta^* \mid \kappa, \omega, \zeta) = \frac{2}{\omega} \times \Phi\left(\zeta \frac{\eta^* - \kappa}{\omega}\right) \times \varphi\left(\frac{\eta^* - \kappa}{\omega}\right), \tag{2}$$

where ω denotes the scale parameter, $\Phi(.)$ denotes the standard normal distribution function, ζ denotes the shape parameter, κ denotes the location parameter and $\varphi(.)$ denotes the standard normal probability density function. Note that the normal distribution is a special case of the skew-normal distribution, with $\zeta = 0$.

The expected value and variance of the distribution are respectively (Azzalini & Capitanio, 1999):

$$E(\eta^*) = \kappa + \omega\sqrt{\frac{2}{\pi}}\frac{\zeta}{\sqrt{1+\zeta^2}}, \tag{3}$$

and

$$Var(\eta^*) = \omega^2\left(1 - \frac{2\zeta^2}{\pi(1+\zeta^2)}\right). \tag{4}$$

To identify the model, the usual scale and metric restrictions have to be imposed. For example, one may take $E(\eta^*) = 0$ and $Var(\eta^*) = 1$. The skew-normal factor model can be fitted by means of Marginal Maximum Likelihood (MML; Bock & Aitkin, 1981; see Molenaar et al., 2010; Montanari & Viroli, 2010).

An extension of the above model is the skew-normal factor model with multiple factors (Q, with $Q \geq 1$; Montanari & Viroli, 2010). This model involves the multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996). That is, the $(Q \times 1)$ vector with factor scores $\boldsymbol{\eta}^* \sim SN(\boldsymbol{\Omega}, \boldsymbol{\alpha})$, with the density function

$$f(\mathbf{t}) = 2\phi_Q(\mathbf{t};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\mathbf{t}), \tag{5}$$

where $\phi_Q(\mathbf{t};\boldsymbol{\Omega})$ is the $Q-$dimensional normal density with mean zero and correlation matrix $\boldsymbol{\Omega}$, and Φ denotes the distribution function of the $N(0,1)$ distribution. The vector $\boldsymbol{\alpha}$ contains the shape parameters (related to ζ in Eq. (2)). The mean vector and covariance matrix of $\boldsymbol{\eta}^*$ are respectively (Montanari & Viroli, 2010):

$$\boldsymbol{\mu}_{\eta^*} = E(\boldsymbol{\eta}^*) = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}, \qquad Var(\boldsymbol{\eta}^*) = \boldsymbol{\Omega} - \boldsymbol{\mu}_{\eta^*}\boldsymbol{\mu}'_{\eta^*}, \tag{6}$$

with $\boldsymbol{\delta} = (1+\boldsymbol{\alpha}'\boldsymbol{\Omega}\boldsymbol{\alpha})^{-1/2}\boldsymbol{\Omega}\boldsymbol{\alpha}$.

The well-known equivalence after orthogonal and oblique transformations holds for the multivariate skew-normal distribution as well. That is, if $\boldsymbol{\eta}^* \sim SN(\boldsymbol{\Omega}, \boldsymbol{\alpha})$, and $\mathbf{H}$ is a non-singular $(Q \times Q)$ matrix such that $\mathbf{H}'\boldsymbol{\Omega}\mathbf{H}$ is a correlation matrix, then $\mathbf{H}'\boldsymbol{\eta}^* \sim SN(\mathbf{H}'\boldsymbol{\Omega}\mathbf{H}, \mathbf{H}^{-1}\boldsymbol{\alpha})$. Azzalini & Capitanio (1999) show that a linear transformation $\mathbf{H}'\boldsymbol{\eta}^*$ exists that transforms the multivariate skew-normal density to a canonical form with $\boldsymbol{\Omega} = \mathbf{I}_Q$ and $\boldsymbol{\alpha}' = \begin{pmatrix}\alpha_1 & 0 & \cdots & 0\end{pmatrix}$. In that case, the first random variable is a unidimensional skew-normal with parameters $(0,\ 1,\ \alpha_1)$, and the other random variables have the $N(0,1)$ distribution. Moreover, the Q random variables are mutually independent since their joint density in (5) equals the product of their marginal densities. In the skew-normal factor model with multiple factors, factors are estimated in this canonical form, in which only a single factor has a shape parameter $\alpha \neq 0$ (Montanari & Viroli, 2010).

5.2.2 The Quadratic Factor Model

The quadratic factor model, a special case of a nonlinear factor model (e.g., McDonald, 1962, 1967; Mooijaart & Bentler, 1986), is just like the skew-normal factor model an approach to account for non-normally distributed expected item scores (Molenaar et al., 2010). Here, we will describe the quadratic factor model.

If y_i denotes a randomly observed score on item *i*, the following nonlinear factor model (Mooijaart & Bentler, 1986) is specified for y_i :

$$y_i = \tilde{\nu}_i + \tilde{\lambda}_i s(\eta) + \varepsilon_i, \tag{7}$$

where $\tilde{\nu}_i$ is the intercept of item *i*, $\tilde{\lambda}_i$ is the factor loading of item *i*, η is the common factor score, $\mathrm{s}(\eta)$ is a function of the factor scores and ε_i is the residual term. The residual and the common factor score are drawn from a normal distribution: $\varepsilon_i \sim N(0,\sigma^2)$ and $\eta \sim N(\mu,\sigma^2)$. For $\mathrm{s}(\eta)$, one may specify a polynomial function as:

$$\mathrm{s}(\eta) = \gamma_{i0} + \gamma_{i1}\eta + \gamma_{i2}\eta^2 + \ldots + \gamma_{ir}\eta^r. \tag{8}$$

In the quadratic factor model, the polynomial function is of degree 2. The model for y_i then becomes:

$$\begin{aligned} y_i &= \tilde{\nu}_i + \tilde{\lambda}_i(\gamma_{i0} + \gamma_{i1}\eta + \gamma_{i2}\eta^2) + \varepsilon_i, \\ y_i &= \tilde{\nu}_i + \tilde{\lambda}_i\gamma_{i0} + \tilde{\lambda}_i\gamma_{i1}\eta + \tilde{\lambda}_i\gamma_{i2}\eta^2 + \varepsilon_i, \\ y_i &= \nu_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2 + \varepsilon_i, \end{aligned} \tag{9}$$

with ν_i the intercept of item *i*, and $\lambda_{i(x)}$ the factor loading associated with the x^{th} power of η for the i^{th} item. This model can be fitted using MML (Bock & Aitkin, 1981; see Molenaar et al., 2010).

5.2.3 Relation between the Skew-Normal Factor Model and the Quadratic Factor Model

Because non-normality of the expected item scores can be modeled with the skew-normal factor model and with the quadratic factor model, the important question arises how these models are related. We will show that the quadratic factor model is equivalent to the skew-normal factor model up to the third-order moments, but that the converse is not generally true. This implies that observed data that follow any skew-normal factor model can be so well approximated with the quadratic factor model that the models are empirically indistinguishable. The reverse does not hold in general.

We consider the skew-normal factor model in (1) and the quadratic factor model in (9). By noting that the residuals ε_i^* and ε_i rely on exactly the same assumptions in both models, and that the distributions of ε_i^* and ε_i are independent from respectively $(y_i - \varepsilon_i^*)$ and $(y_i - \varepsilon_i)$, we can leave the residual variances aside in comparing the models. It remains to address the differences in distributions of the expected item scores of the two models.

Under the skew-normal factor model, T_i^*, the expected score on item *i*, equals

$$T_i^* = \nu_i^* + \lambda_i^* \eta^*, \tag{10}$$

where ν_i^* is the intercept of item *i*, λ_i^* is the factor loading of item *i*, and η^* is the common factor score. The common factor score η^* is drawn from a skew-normal distribution: $\eta^* \sim SN(\kappa,\omega,\zeta)$, with location parameter κ, scale parameter ω, and shape parameter ζ. As it stands, the model is unidentified; to identify the model, we fix κ at 0 and ω at 1, so that $T_i^* = \nu_i^* + \lambda_i^* \eta^*$, with $\eta^* \sim SN(0,1,\zeta)$.

Under the quadratic factor model, T_i, the expected score on item *i*, equals

$$T_i = \nu_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2, \tag{11}$$

where ν_i is the intercept of item *i*, $\lambda_{i(1)}$ and $\lambda_{i(2)}$ are the factor loadings of item *i*, and η is the common factor score from a normal distribution: $\eta \sim N(\mu,\sigma^2)$. To identify the model, we fix μ at 0 and σ^2 at 1, so that $T_i = \nu_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2$, with $\eta \sim N(0,1)$. To have fully equivalent models, the densities of T_i and T_i^* should be equal. This is not generally true. We will first consider approximating the density of T_i^* (i.e., under the skew-factor model) by that of T_i (i.e., under the quadratic factor model), and then the converse.

Approximating the skew-normal factor model by a quadratic factor model

As we will show, for suitably chosen parameter values, the density of T_i, the distribution of the expected scores under a quadratic factor model, is to its third moment equivalent to the density of T_i^*, the distribution of the expected scores under the skew-normal factor model. For simplification, we consider a special case of the skew-normal factor model for T_i^*, by fixing ν_i^* at 0 and λ_i^* at 1, so that $T_i^* = \eta^*$. This can be done without loss of generality, because any differences in location and scale of T_i and T_i^* can be solved through ν_i, $\lambda_{i(1)}$ and $\lambda_{i(2)}$.

The question now reduces to whether there exists constants ν_i, $\lambda_{i(1)}$ and $\lambda_{i(2)}$ which equate the first three moments of the density of $T_i = \nu_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2$ (Eq. 11), to the density of η^*, with $\eta^* \sim SN(\kappa,\omega,\zeta)$. In Appendix A, it is proven that those constants exist (omitting the index *i* in ν_i, $\lambda_{i(1)}$, $\lambda_{i(2)}$ and T_i to improve the readability). To find the constants ν_i, $\lambda_{i(1)}$ and $\lambda_{i(2)}$, one first needs to find $\lambda_{i(2)}$ (by solving Eq. (A6) in Appendix A for λ_2 that satisfies $1 - c_1^2 - 2\lambda_{(2)}^2 > 0$). Then, $\lambda_{i(1)}$ and ν_i can be computed as:

$$\lambda_{i(1)} = \sqrt{1 - c_1^2 - 2\lambda_{i(2)}^2}, \qquad \nu_i = c_1 - \lambda_{i(2)}.$$

Further, it is shown in Appendix A that $\lambda_{i(2)}$ and ν_i are unique, while $\lambda_{i(1)}$ is unique up to sign. The latter is because the distribution of the term $\lambda_{i(1)}\eta$ in T_i is symmetric around zero, and hence does not depend on the sign of $\lambda_{i(1)}$.

The preceding implies that the density of T_i^*, the expected scores under the skew-normal factor model, can be approximated by the density of T_i, the expected scores under the quadratic factor model. To illustrate how well the density of T_i approximates the density of $T_i^* = \eta^*$, we consider the cases with $\zeta = 1.81$, $\zeta = 2.17$,

$\zeta = 2.62$, and $\zeta = 3.50$, corresponding to a small, medium, large and very large coefficient, respectively (see Molenaar et al., 2010, for an indication of the magnitude of shape parameters). To assess the closeness of the densities of η^* and T_i, we use the L_1-norm of the difference between their densities:

$$\left\| f_{\eta^*} - f_{T_i} \right\|_1 = \int \left| f_{\eta^*}(y) - f_{T_i}(y) \right| dy.$$

The L_1-norm of the density differences and the values for ν_i, $\lambda_{i(1)}$, and $\lambda_{i(2)}$ can be found in Table 5.1. As can be seen, the L_1-norm is rather small, even for large values of ζ. Further, the closeness of the distributions of η^* and T_i is illustrated in Figure 5.1. Here, the densities of η^* and T_i are plotted for $\zeta = 1.81$ (Figure 5.1a), $\zeta = 2.17$ (Figure 5.1b), $\zeta = 2.62$ (Figure 5.1c), and $\zeta = 3.50$ (Figure 5.1d). As can be seen, they are very close. This indicates that a linear model with a skew-normal factor can be well approximated by a quadratic factor model with a normal factor. This implies that in practice, the two models are empirically indistinguishable from each other. This also explains that the quadratic factor model cannot be jointly implemented with a skew-normal factor, as demonstrated by Molenaar et al. (2010).

Above, we illustrated how well T_i approximates η^*, where T_i are the expected scores under the quadratic factor model, a nonlinear factor model with a polynomial of the second degree. We conjecture that when the degree of the polynomial would be taken larger than two, the approximation of η^*, by T_i will improve. This can be expected because the skew-normal distribution is completely determined by its moments (Gupta, Nguyen, & Sanqui, 2004; Lemma 2.1). As a result, if the moments $E(T^k)$ converge to the moments of $E(\eta^{*k})$ for k = 1, 2, …, then the distribution of T_i will converge to the distribution of η^* (e.g., Billingsley, 1995; section 30). Therefore, taking the degree of the polynomial T_i larger than two, more moments could be equated, and if equality holds for more moments $E(T^k) = E(\eta^{*k})$, then the closeness of the distributions η^* and T_i will be even better than we already had with a second degree polynomial.

Table 5.1. The L_1-norm of the Density Differences of η^* and T_i, and the values for ν_i, $\lambda_{i(1)}$, and $\lambda_{i(2)}$, for different values of ζ.

	$\zeta = 1.81$	$\zeta = 2.17$	$\zeta = 2.62$	$\zeta = 3.50$
L_1-norm	0.0271	0.0360	0.0468	0.0669
ν_i	0.6507	0.6671	0.6783	0.6879
$\lambda_{i(1)}$	0.7125	0.6843	0.6598	0.6315
$\lambda_{i(2)}$	0.0477	0.0576	0.0671	0.0793

The converse: Approximating the quadratic factor model by a skew-normal factor model

The question is whether for any given constants ν_i, $\lambda_{i(1)}$, and $\lambda_{i(2)}$, there exists values of ν_i^*, λ_i^* and ζ for which the first three moments of the density of $T_i^* = \nu_i^* + \lambda_i^* \eta^*$, with $\eta^* \sim SN(0,1,\zeta)$ (Eq. 10) could be made equal to those of the density of $T_i = \nu_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2$, with $\eta \sim N(0,1)$ (Eq. 11). In Appendix B, it is proven that this cannot be done in all cases.

The limiting factor in equating the densities appears to be the limited range in skewness (and kurtosis) of the skew-normal distribution (see e.g., Azzalini, 1985; 2005; Henze, 1986). As a result, for large values of $\lambda_{i(2)}$, the skewness of T_i is outside the range of the skewness of T_i^*. Consequently, for these values of $\lambda_{i(2)}$, the first three moments of the density of T_i^* cannot be made equal to those of the density of T_i. In Appendix C it is shown that for small values of $\lambda_{i(2)}$ (roughly between -0.17 and 0.17), for which the skewness of T_i fall within the range of the skewness of T_i^*, the first three moments of T_i^* can be equated to the first three moments of T_i.

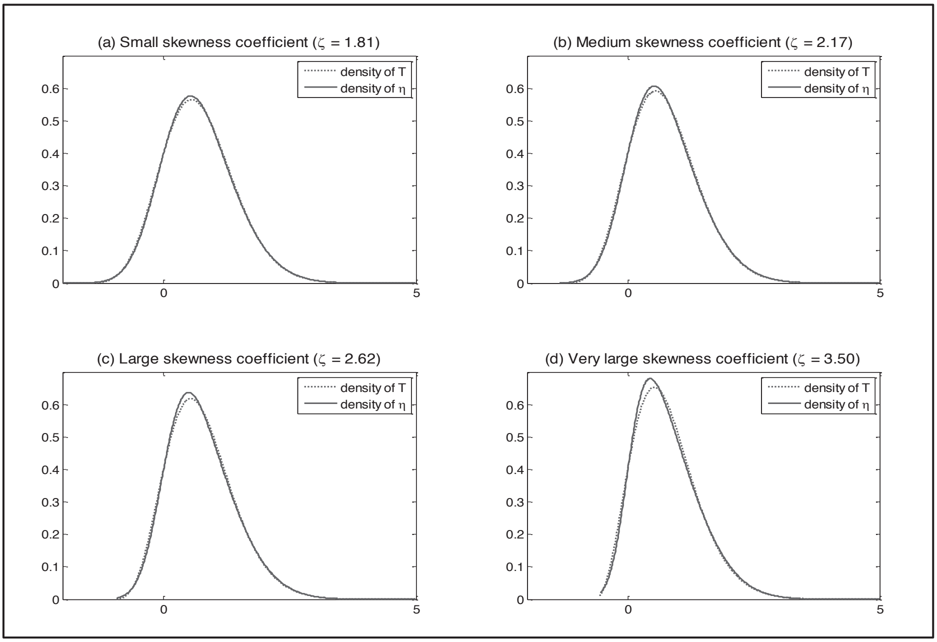

Figure 5.1. Closeness of the Distributions of η^* and the approximation T for a small ($\zeta = 1.81$), medium ($\zeta = 2.17$), large ($\zeta = 2.62$), and very large ($\zeta = 3.50$) skewness coefficient.

The multiple factor case

As we showed above, a skew-normal factor from a skew-normal factor model for a single factor can be very well approximated by a quadratic factor model, and vice versa, if the loading of the quadratic function is within certain bounds. In Appendix D, it is shown that this relation between the skew-normal factor model for a single factor and the nonlinear factor model can be generalized to the case with multiple factors.

That is, if the expected score on item i in the skew-normal factor model is $T_i^* = \nu_i^* + \lambda_{i(1)}^* \eta_1^* + \sum_{q=2}^{Q} \lambda_{i(q)}^* \eta_q^*$, where the Q–dimensional skew-normal distribution is in canonical form (with η_1^* skew-normal and $\eta_q^* \sim N(0,1)$, $q = 2,\cdots,Q$), then we can find parameters ν_i, $\lambda_{i(11)}$ and $\lambda_{i(12)}$, $\lambda_{i(q)}$, $q = 2,\cdots,Q$, such that the nonlinear factor model with expected item score $T_i = \nu_i + \lambda_{i(11)}\eta_1 + \lambda_{i(12)}\eta_1^2 + \sum_{q=2}^{Q}\lambda_{i(q)}\eta_q$ (with η_q, $q = 1,\cdots,Q$ independent $N(0,1)$ variables) satisfies $E(T_i^{*k}) = E(T_i^k)$, for $k = 1, 2, 3$.

This generalization of the relation between the nonlinear factor model and the skew-normal factor model for a single factor to the case with multiple factors holds because in the canonical skew-normal factor model with multiple factors, all factors are mutually independent, and only a single factor has a shape parameter $\alpha \neq 0$ (Montanari & Viroli, 2010). This implies that analogous to the single factor case, a skew-normal factor model with multiple factors can be well approximated by a nonlinear factor model, and conversely if the skewness of T_i does not fall outside the range of the skewness of T_i^*.

5.2.4 Implications of the Relation between the Skew-Normal Factor and Quadratic Factor Model

As we showed, a skew-normal factor model can be very well approximated by a quadratic factor model, and vice versa, if the loading of the quadratic function is within certain bounds. From a mathematical point of view, in those conditions the choice between these parameterizations is arbitrary. In empirical practice, a skew-normal factor model may be preferred over the quadratic factor model, since one needs fewer parameters, which yields more efficient estimates. Moreover, one uses linear relations between the items and the latent trait, which are generally easier to interpret than nonlinear relations.

5.3 Modeling Non-Normal Expected Item Scores in Multiple Groups

For a single group, the choice between a skew-normal factor model and the quadratic factor model is not crucial, when the data follow a skew-normal factor model or when the loading of the quadratic part is within certain bounds. In contrast, for multiple groups, the difference appears to be crucial when establishing measurement invariance.

5.3.1 Establishing Measurement Invariance

Measurement invariance (MI) holds if a test measures the same latent trait equally across groups. That is, conditional on the level of the latent trait, the expected scores of the items of the test should be group independent (Mellenbergh, 1985, 1989). More formally, items of a test are measurement invariant with respect to group *G*, given the latent trait *Z* (i.e., factor), "if and only if:

$$f(X \mid g,z) = f(X \mid z),$$

for all values *g* and *z* of the variables *G* and *Z*, where $f(X \mid g,z)$ is the distribution of the item response given *g* and *z* and $f(X \mid z)$ the distribution of the item responses given *z*; otherwise the item is biased" (Mellenbergh, 1989, p.129). Here, $f(X \mid z)$ can be any function, and *Z* can have any distribution.

Various statistical methods have been developed to assess whether measurement invariance holds (e.g., Millsap & Everson, 1991). Typically, measurement invariance is examined by means of a multi-group confirmatory factor analysis (MG-CFA) where the latent trait *Z* is assumed to follow a normal distribution and it is common to specify a linear function for $f(X \mid z)$ using a linear factor model for continuous items and an ordinal FA model / GRM for ordinal items.

The requirements for measurement invariance in a linear one factor model for continuous items can be described as follows. If y_{ig} denotes an observed score of a random subject in group *g* on item *i*, we specify the following linear factor model for y_{ig}:

$$y_{ig} = \nu_{ig} + \lambda_{ig}\eta_g + \varepsilon_{ig}, \qquad (12)$$

where ν_{ig} is the intercept of item *i* in group *g*, λ_{ig} is the factor loading of item *i* in group *g*, η_g is the common factor score of a random subject in group *g* (*g* = 1, … *G*) and ε_{ig} is the residual term. An item is strict measurement invariant if all measurement parameters are equal across groups, thus if the intercept, factor loading and the residual term of the item are equal across the *G* groups (i.e., if: $\nu_{ig} = \nu_i$, $\lambda_{ig} = \lambda_i$ and $\varepsilon_{ig} = \varepsilon_i$; see e.g., Meredith, 1993). In an ordinal FA model / GRM analogous requirements apply, with additionally equal relationships between the observed ordinal items and the latent continuous items (see e.g., Millsap & Yun-Tein, 2004).

Tests for measurement invariance (MI) crucially depend on a correct model specification. That is, if a linear factor model is applied to data that follow a quadratic factor model, then this yields diverging loadings and intercepts across groups. This is so because factor mean differences increase between groups (Bauer, 2005). MI tests can be heavily affected by such model misspecifications. As shown in a simulation study, MI tests increasingly rejected MI incorrectly when factor mean differences increased (Bauer, 2005); this stresses the importance of a correct model specification in tests for MI.

In addition, tests for MI can be affected by violations of the normality assumption of the latent trait. As shown in a simulation study of Woods (2008), MI tests too often incorrectly rejected MI[2] when under a two-parameter logistic model (2PLM) the latent trait densities differed across groups. In particular, when the latent trait density was skewly distributed in one group and normally distributed in the other group, MI tests were biased (Woods, 2008). Because the 2PLM for binary items is closely related to the GRM for ordinal items, one may expect those tendencies to hold under the ordinal FA model / GRM as well.

To remedy the biased MI tests, we propose to include a skewness parameter in the models when assessing measurement invariance. That is, to assess measurement invariance with a linear measurement model (i.e., a linear FA model or an ordinal FA model / GRM) that allows for a skew-normal factor score distribution: $\eta_g \sim SN(\kappa_g, \omega_g, \zeta_g)$, with κ_g, ω_g, and ζ_g respectively the location parameter, scale parameter and shape parameter in group g. The key difference with the currently standard approach is that now ζ_g, is not fixed to be equal to zero, but is allowed to differ from zero in all models, across all groups.

MI assessment with a skew-normal factor model can be seen as an alternative to considering nonlinear factors. When data in two or more groups follow a nonlinear factor model and the quadratic term is sufficiently small, the skew-normal factor model offers a very good approximation. In this case, the skew-normal factor model

[2] Woods (2008) uses the term Differential Item Functioning, which is in the IRT tradition the term for violation of MI.

appears to be preferred in view of the fewer parameters. Further, if data in two or more groups follow a skew-normal factor model, with different sizes of the skewness ($\zeta_1 \neq \zeta_2$), MI assessment with a nonlinear factor model appears to be unsuitable. That is, although the skew-normal factor model can be well approximated with a nonlinear factor model, and thus the skewness of the latent trait can be accounted for by nonlinear measurement parameters, they will do so differently across groups because the skewness differs across groups. As a consequence, differences between groups in a structural parameter of the model (i.e., differences in skewness of the latent trait) will in a nonlinear factor model be translated to differences between groups in a measurement parameter of the model (i.e., differences in factor loadings of the quadratic term). This could suggest that MI is violated, whereas in fact it is just the distribution of the latent trait that differs across the groups. Therefore, we generally prefer to use a skew-normal factor model for the assessment of MI across groups.

5.4 Discussion

Deviations from normality of the expected item scores can be modeled through either a skew-normal factor model or through a quadratic factor model (see Molenaar et al., 2010). In this chapter, we showed why these two variants to account for non-normal expected item scores cannot be implemented jointly into a single model. We showed that the quadratic factor model is equivalent to the skew-normal factor model up to the third-order moments, and that the converse is true if the factor loading of the quadratic term is small. Furthermore, we show that the intimate relation between the skew-normal factor model and the quadratic factor model holds for both the single and multiple factor case. We illustrated that observed data that follow any skew-normal factor model can be so well approximated with the quadratic factor model that the models are indistinguishable in practice. In a single-group context, the choice between the two models is not crucial. In contrast, in establishing measurement invariance across populations, it can be of key importance to explicitly model the skewness of the factor distribution. This can be done using a model that incorporates a skew-normal factor (Molenaar et al., 2010, 2012; Molenaar, Dolan, & van der Maas, 2011; Montanari & Viroli, 2010; for a Bayesian approach see Azevedo, Andrade, & Fox, 2012) or a model that approximates the factor distribution. The latter can be done nonparametrically (Woods, 2011a, 2001b, 2011c), or semi-nonparametrically (Irincheeva, Cantoni, & Genton, 2012).

We note that deviations from normality go beyond its skewness, and are aware of the fact that higher moments of a distribution such as the kurtosis may be of importance as well in the comparison of multiple populations. However, we consider the first three moments as a good start in comparing multiple populations and advice researchers to establish measurement invariance in a model in which a parameter for the mean, variance and skewness of the latent trait is considered.

5.5 Appendix A

Proposition 1. *Let* η^* *have the skew-normal distribution with parameters* $(0, 1, \zeta)$. *Let* $T = \nu + \lambda_{(1)}\eta + \lambda_{(2)}\eta^2$, *where* η *has the* $N(0,1)$ *distribution, and* ν, $\lambda_{(1)}$, $\lambda_{(2)}$ *are real constants. Then, for any* ζ, *there exist constants* ν, $\lambda_{(1)}$, $\lambda_{(2)}$ *such that* $E(\eta^{*k}) = E(T^k)$ *for k* = 1, 2, 3.

Proof. The moments of the skew normal distribution can be found in Corollary 4 of Henze (1986). For $k = 1$, we have

$$E(\eta^*) = \sqrt{\frac{2}{\pi}}\frac{\zeta}{\sqrt{1+\zeta^2}} = c_1, \qquad E(T) = \nu + \lambda_{(2)}.$$

Hence, we obtain

$$\nu = c_1 - \lambda_{(2)}. \tag{A1}$$

For $k = 2$, we have

$$E(\eta^{*2}) = 1, \qquad E(T^2) = \nu^2 + 2\nu\lambda_{(2)} + 3\lambda_{(2)}^2 + \lambda_{(1)}^2.$$

Hence, we obtain

$$\nu^2 + 2\nu\lambda_{(2)} + 3\lambda_{(2)}^2 + \lambda_{(1)}^2 - 1 = 0. \tag{A2}$$

For $k = 3$, we have

$$E(\eta^{*3}) = \sqrt{\frac{2}{\pi}}\frac{3\zeta}{(1+\zeta^2)^{3/2}}(1+\frac{4\zeta^2}{6}) = c_3, \qquad E(T^3) = \nu^3 + 3\nu^2\lambda_{(2)} + 3\nu\lambda_{(1)}^2 + 9\nu\lambda_{(2)}^2 + 9\lambda_{(1)}^2\lambda_{(2)} + 15\lambda_{(2)}^3.$$

Hence, we obtain

$$\nu^3 + 3\nu^2\lambda_{(2)} + 3\nu\lambda_{(1)}^2 + 9\nu\lambda_{(2)}^2 + 9\lambda_{(1)}^2\lambda_{(2)} + 15\lambda_{(2)}^3 - c_3 = 0. \tag{A3}$$

Next, we substitute the expression (A1) for ν into (A2) and (A3). After simplifying, we obtain the following two equations:

$$2\lambda_{(2)}^2 + \lambda_{(1)}^2 + c_1^2 - 1 = 0, \tag{A4}$$

$$8\lambda_{(2)}^3 + 6c_1\lambda_{(2)}^2 + 6\lambda_{(1)}^2\lambda_{(2)} + 3c_1\lambda_{(1)}^2 + c_1^3 - c_3 = 0. \tag{A5}$$

Note that (A4) only has a real solution for $\lambda_{(1)}, \lambda_{(2)}$ if $c_1^2 < 1$. This holds for all ζ, because it is equivalent to $\zeta^2/(1+\zeta^2) < \pi/2$.

Next, we rewrite (A4) as $\lambda_{(1)}^2 = 1 - c_1^2 - 2\lambda_{(2)}^2$ and substitute this for $\lambda_{(1)}^2$ in (A5). After simplifying, this yields

$$-4\lambda_{(2)}^3 + 6(1-c_1^2)\lambda_{(2)} - 2c_1^3 + 3c_1 - c_3 = 0. \tag{A6}$$

We show that this third degree polynomial in $\lambda_{(2)}$ has three distinct real roots for any shape parameter ζ, of which only one root satisfies $1 - c_1^2 - 2\lambda_{(2)}^2 > 0$ (see below). Then, $\lambda_{(1)}$ and ν can be computed as:

$$\lambda_{(1)} = \sqrt{1 - c_1^2 - 2\lambda_{(2)}^2}, \qquad \nu = c_1 - \lambda_{(2)}.$$

Here, both $\lambda_{(2)}$ and ν are unique, while $\lambda_{(1)}$ is unique up to sign. The latter is because the distribution of the term $\lambda_{(1)}\eta$ in T does not depend on the sign of $\lambda_{(1)}$ (it is symmetric around zero).

It remains to show that for any shape parameter ζ, the third degree polynomial (A6) in $\lambda_{(2)}$ has three distinct real roots, and that exactly one root satisfies $1 - c_1^2 - 2\lambda_{(2)}^2 > 0$.

The discriminant of a general third degree polynomial $ax^3 + bx^2 + cx + d$ is defined as

$$D = 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2.$$

The polynomial has three distinct roots if and only if $D > 0$, see e.g. section 10.3 of Irving (2004). For the polynomial (A6), the discriminant depends on ζ. We have

$$D(\zeta) = 16 \cdot 6^3(1-c_1^2)^3 - 27 \cdot 16(-2c_1^3 + 3c_1 - c_3)^2.$$

Using symbolic computation software, it can be verified that $\pi^3(1+\zeta^2)^3 D(\zeta)$ equals

$$3456\pi^3 + 10368(-2\pi^2 + \pi^3)\zeta^2 + 10368(4\pi - 4\pi^2 + \pi^3)\zeta^4 + 864(-48 + 56\pi - 25\pi^2 + 4\pi^3)\zeta^6.$$

This sixth degree polynomial in ζ has six complex roots. It follows that $D(\zeta) > 0$ for any ζ if and only if $D(\zeta) > 0$ for some ζ. Since $D(0) = 3456,$ we have proven that the polynomial (A6) has three distinct real roots for any ζ.

Setting the derivative of (A6) to zero yields

$$-12\lambda_{(2)}^2 + 6(1 - c_1^2) = 0.$$

Hence, the local minimum and local maximum of (A6) are found at

$$\lambda_{(2)}^{(\min)} = -\sqrt{(1 - c_1^2)/2}, \qquad \lambda_{(2)}^{(\max)} = \sqrt{(1 - c_1^2)/2}.$$

Note that $1 - c_1^2 > 0$ for any ζ, as shown below (A5). Also note that the coefficient of $\lambda_{(2)}^3$ in (A6) is negative, which implies that $\lambda_{(2)}^{(\min)} < \lambda_{(2)}^{(\max)}$.

Since the polynomial (A6) has three real roots, there is exactly one root in between $\lambda_{(2)}^{(\min)}$ and $\lambda_{(2)}^{(\max)}$. We have

$$1 - c_1^2 - 2(\lambda_{(2)}^{(\min)})^2 \ = \ 1 - c_1^2 - 2(\lambda_{(2)}^{(\max)})^2 \ = \ 0.$$

Hence, for the root $\lambda_{(2)}^{*}$ in between $\lambda_{(2)}^{(\min)}$ and $\lambda_{(2)}^{(\max)}$ it holds that $1 - c_1^2 - 2(\lambda_{(2)}^{*})^2 > \ 0.$ This completes the proof.

5.6 Appendix B

Here, we show that the converse of Proposition 1 is not true. That is, for some constants ν, $\lambda_{(1)}$, $\lambda_{(2)}$, there do not exist values for ν^*, λ^* and ζ that equate the first three moments of $T^* = \nu^* + \lambda^*\eta^*$ and $T = \nu + \lambda_{(1)}\eta + \lambda_{(2)}\eta^2$.

For simplicity, we set $\nu = 0$ and $\lambda_{(1)} = 1$. Note that T^* has a skew-normal distribution with parameters $(\nu^*, \lambda^*, \zeta)$. If equating the first three moments of T^* and T would be possible, then also their skewnesses would be equal. That is,

$$E\left(\frac{T - E(T)}{\sqrt{Var(T)}}\right)^3 = E\left(\frac{T^* - E(T^*)}{\sqrt{Var(T^*)}}\right)^3.$$

For the left-hand side, we compute

$$E\left(\frac{T - E(T)}{\sqrt{Var(T)}}\right)^3 = \frac{E\left(T - E(T)\right)^3}{\left(E\left(T - E(T)\right)^2\right)^{3/2}} = \frac{8\lambda_{(2)}^3 + 6\lambda_{(2)}}{\left(2\lambda_{(2)}^2 + 1\right)^{3/2}}. \tag{B1}$$

The skewness of T^* only depends on ζ. From Azzalini (1985) we obtain

$$E\left(\frac{T^* - E(T^*)}{\sqrt{Var(T^*)}}\right)^3 = \left(\frac{4-\pi}{2}\right)\frac{\left(\sqrt{\frac{2}{\pi}}\frac{\zeta}{\sqrt{1+\zeta^2}}\right)^3}{\left(1 - \frac{2}{\pi}\left(\frac{\zeta^2}{1+\zeta^2}\right)\right)^{3/2}}. \tag{B2}$$

As $\left|\lambda_{(2)}\right|$ becomes very large, it can be seen that the skewness of T in (B1) converges to

$$\pm\sqrt{8} = \pm 2\sqrt{2} \approx \pm 2.82.$$

As $|\zeta|$ becomes very large, the skewness of T^* in (B2) converges to

$$\pm\left(\frac{4-\pi}{2}\right)\left(\frac{2/\pi}{1-2/\pi}\right)^{3/2} \approx \pm 0.9953.$$

For large values of $\left|\lambda_{(2)}\right|$ the skewness of T is outside the range of the skewness of T^{*}. We therefore conclude that the converse statement of Proposition 1 does not hold.

5.7 Appendix C

Here, we show that if $\lambda_{(2)}$ is small enough, such that the skewness of T is not outside the range of the skewness of T^*, that then the first three moments of T^* can be equated to the first three moments of T. Note that we may set $\nu = 0$ and $\lambda_{(1)} = 1$ without loss of generality, since the location and scaling can be absorbed in the parameters ν^* and λ^*.

Proposition 2. *Let $T = \eta + \lambda_{(2)}\eta^2$, where η has the $N(0,1)$ distribution, and $\lambda_{(2)}$ is a real constant such that*

$$\left| \frac{8\lambda_{(2)}^3 + 6\lambda_{(2)}}{\left(2\lambda_{(2)}^2 + 1\right)^{3/2}} \right| \leq \left(\frac{4-\pi}{2}\right)\left(\frac{2/\pi}{1-2/\pi}\right)^{3/2}. \qquad \text{(C1)}$$

Let T^* have the skew-normal distribution with parameters $(\nu^*, \lambda^*, \zeta)$. Then, there exist parameters ν^*, λ^*, ζ such that $E(T^{*k}) = E(T^k)$ for k = 1, 2, 3.

Proof. Using Azzalini (1985) for the moments of the skew-normal distribution, we obtain

$$E(T) = \lambda_{(2)}, \qquad E(T^*) = \nu^* + \lambda^* \sqrt{\frac{2}{\pi}} \frac{\zeta}{\sqrt{1+\zeta^2}}, \qquad \text{(C2)}$$

$$E(T^2) = 3\lambda_{(2)}^2 + 1, \qquad E(T^{*2}) = \nu^* + 2\nu^*\lambda^* \sqrt{\frac{2}{\pi}} \frac{\zeta}{\sqrt{1+\zeta^2}} + \lambda^{*2}. \qquad \text{(C3)}$$

In Appendix B above, the skewness of T is given in (B1) and the skewness of T^* in (B2). Since the skewness of T^* in (B2) depends only on ζ, we estimate ζ by equating the skewnesses of T and T^*. This is possible by the requirement (C1). Let the skewness of T in (B1) be denoted by γ. We substitute $\delta = \zeta / \sqrt{1+\zeta^2}$.

Setting (B2) equal to γ and solving for δ yields

$$|\delta| = \sqrt{\frac{(\pi/2)|\gamma|^{2/3}}{|\gamma|^{2/3} + (2-\pi/2)^{2/3}}}, \tag{C4}$$

with δ and γ having the same sign. Next, we obtain ζ as $\zeta = \delta / \sqrt{1-\delta^2}$.

When ζ is known, we equate the first and second moments of T and T^* to obtain ν^* and λ^*. Since the skewnesses of T and T^* are equal, it then follows that also $E(T^3) = E(T^{*3})$.

Setting $E(T) = E(T^*)$ in (C2) yields

$$\nu^* = \lambda_{(2)} - \sqrt{\frac{2}{\pi}}\delta\lambda^*. \tag{C5}$$

Setting $E(T^2) = E(T^{*2})$ in (C3) and substituting (C5) for ν^* yields after rewriting

$$2\lambda_{(2)}^2 + 1 = \lambda^{*2}\left(1 - \frac{2\delta^2}{\pi}\right).$$

Since λ^* is the scaling parameter of a skew-normal distribution, it must be positive. Hence, we obtain

$$\lambda^* = \sqrt{\frac{2\lambda_{(2)}^2 + 1}{1 - \frac{2\delta^2}{\pi}}}. \tag{C6}$$

After λ^* is known, we obtain ν^* from (C5). This completes the proof.

5.8 Appendix D

Here, we will show that the demonstrated relation between the skew-normal factor model and the quadratic factor model generalizes to the multiple factor case.

Proposition 3. *Let* $\boldsymbol{\eta}^*$ *have a Q dimensional skew-normal distribution, defined by (5) and (6), in the canonical form with* $\boldsymbol{\Omega}=\mathbf{I}_Q$ *and* $\boldsymbol{\alpha}'=\begin{pmatrix}\alpha_1 & 0 & \cdots & 0\end{pmatrix}$. *Let*

$$T^* = \nu^* + \lambda_1^*\eta_1^* + \sum_{q=2}^{Q}\lambda_q^*\eta_q^*.$$

Let

$$T = \nu + \lambda_{11}\eta_1 + \lambda_{12}\eta_1^2 + \sum_{q=2}^{Q}\lambda_q\eta_q,$$

where $\eta_1,\ldots\eta_Q$ *are mutually independent* $N(0,1)$ *variables. Then, for any* α_1, ν^*, $\lambda_1^*,\ldots,\lambda_Q^*$ *there exist constants* ν, λ_{11}, λ_{12}, $\lambda_2,\ldots,\lambda_Q$ *such that* $E(T^{*k}) = E(T^k)$ *for* $k = 1, 2, 3$.

Proof. Without loss of generality we set $\nu^* = 0$ and $\lambda_1^* = 1$. Note that η_1^* has a unidimensional skew-normal distribution with parameters $(0, 1, \alpha_1)$, and η_q^*, $q = 2, \cdots, Q$, are $N(0,1)$ distributed. Moreover, $\eta_1^*, \ldots, \eta_Q^*$ are mutually independent. Let $T_0^* = \eta_1^*$ and $T_0 = \nu + \lambda_{11}\eta_1 + \lambda_{12}\eta_1^2$. From Proposition 1 we know that for any α_1 there exist ν, λ_{11}, λ_{12}, such that $E(T_0^{*k}) = E(T_0^k)$ for $k = 1, 2, 3$. Let ν, λ_{11}, λ_{12} have these values. Then $E(T^*) = E(T_0^*) = E(T_0) = E(T)$ holds.

We have

$$\begin{aligned} E(T^{*2}) &= E(T_0^{*2} + 2T_0^*(T^* - T_0^*) + (T^* - T_0^*)^2) \\ &= E(T_0^{*2}) + 2E(T_0^*)E(T^* - T_0^*) + E(T^* - T_0^*)^2 \\ &= E(T_0^{2}) + \sum_{q=2}^{Q}\lambda_q^{*2}, \end{aligned}$$

where we used the independence of T_0^* and $T^* - T_0^*$ in the second step, and $E(T^* - T_0^*) = 0$ and $E(T_0^{*2}) = E(T_0^2)$ in the third step. We set $\lambda_q = \lambda_q^*$ for $q = 2, \cdots, Q$. Then an analogous expansion of $E(T^2)$ shows that $E(T^{*2}) = E(T^2)$.

We have

$$
\begin{aligned}
E(T^{*3}) &= E(T_0^{*3} + 3T_0^{*2}(T^* - T_0^*) + 3T_0^*(T^* - T_0^*)^2 + (T^* - T_0^*)^3) \\
&= E(T_0^{*3}) + 3E(T_0^{*2})E(T^* - T_0^*) + 3E(T_0^*)E(T^* - T_0^*)^2 + E(T^* - T_0^*)^3 \\
&= E(T_0^3) + 3E(T_0)\left(\sum_{q=2}^{Q} \lambda_q^{*2}\right),
\end{aligned}
$$

where we used the independence of T_0^* and $T^* - T_0^*$ in the second step, and $E(T^* - T_0^*) = 0$, $E(T^* - T_0^*)^3 = 0$, $E(T_0^{*3}) = E(T_0^3)$ and $E(T_0^*) = E(T_0)$ in the third step. As above, an analogous expansion of $E(T^3)$ shows that $E(T^{*3}) = E(T^3)$. This completes the proof.

As in the univariate case, the full converse result of Proposition 3 does not hold but a partial converse result is possible under a condition on the skewness of T. This result is omitted here.