



Modelling non-normal data: The relationship between the skew-normal factor model and the quadratic factor model

Iris A. M. Smits*, Marieke E. Timmerman and Alwin Stegeman

Heymans Institute for Psychological Research, Department of Psychometrics and Statistics, University of Groningen, The Netherlands

Maximum likelihood estimation of the linear factor model for continuous items assumes normally distributed item scores. We consider deviations from normality by means of a skew-normally distributed factor model or a quadratic factor model. We show that the item distributions under a skew-normal factor are equivalent to those under a quadratic model up to third-order moments. The reverse only holds if the quadratic loadings are equal to each other and within certain bounds. We illustrate that observed data which follow any skew-normal factor model can be so well approximated with the quadratic factor model that the models are empirically indistinguishable, and that the reverse does not hold in general. The choice between the two models to account for deviations of normality is illustrated by an empirical example from clinical psychology.

I. Introduction

Maximum likelihood estimation of the linear factor model for continuous items is based on the assumption of normally distributed item scores. Non-normality of item scores occurs in empirical practice, giving a need for alternatives. 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 non-parametric maximum likelihood estimation (Skrondal & Rabe-Hesketh, 2004, pp. 182–184). Parametric alternatives are, for the case with latent exogenous variables, the latent moderated structural equations approach (Klein & Moosbrugger, 2000), and, for the case without such exogenous variables, the structural equation finite mixture model (Jedidi, Jagpal, & DeSarbo, 1997) and the non-linear factor model (Mooijaart & Bentler, 1986). Non-normality of the residuals can be modelled as a function of the latent trait score, thereby allowing for heteroscedasticity of the residuals (Hessen & Dolan, 2009). Recently, the skew-normal factor model (Molenaar, Dolan, & Verhelst, 2010; Montanari & Viroli, 2010) has been proposed, which pertains to a linear factor model with skew-distributed factors.

The parametric approaches to account for non-normality of continuous item scores have been put into a single framework (Molenaar *et al.*, 2010). Of particular interest are the two variants that account for deviations from normality of the conditional

*Correspondence should be addressed to Iris A. M. Smits, Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands (email: i.a.m.smits@uva.nl).

means (i.e., the expected item scores conditioned upon the factor). This is done either through a quadratic factor model, a non-linear factor model with a polynomial of the second degree, or through a skew-normal factor model. On the basis of empirical identification checks, Molenaar *et al.* (2010) illustrated that the latter two variants cannot be implemented jointly in 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?

We note that maximum likelihood estimation of the linear factor model is asymptotically robust to non-normality (Amemiya & Anderson, 1990; Anderson & Amemiya, 1988). Further, fitting a linear factor model under normality assumptions to observed data that follow a skew-normal factor model yields consistent estimates (Shapiro, 1984). In limited sample sizes, explicitly modelling the skewness may improve the estimates, giving a need for the skew-normal factor model in empirical practice. When observed data follow a quadratic factor model, parameter estimates of the linear factor model are biased (Bauer, 2005), rendering the use of the quadratic model necessary for this type of data.

In this study, we consider and analyse the relationship between the quadratic factor model and the skew-normal factor model as variants to model non-normal conditional means, for both a single factor and multiple factors. That is, we show that in the case of a skew-normal factor the two variants are empirically indistinguishable, whereas the reverse does not generally hold. We discuss the implications of this relationship for the choice between the variants, and illustrate with an empirical example how one may model deviations from normality.

2. Modelling non-normal conditional means

2.1. The skew-normal factor model

The skew-normal factor model has been proposed to account for non-normally distributed conditional means. Molenaar *et al.* (2010) proposed the model for a single factor, and, independently, Montanari and Viroli (2010) proposed the general model, involving multiple factors.

We start by introducing 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:

$$y_i = v_i^* + \lambda_i^* \eta^* + \varepsilon_i^*, \quad (1)$$

where v_i^* is the intercept of item i , λ_i^* the factor loading, η^* the common factor score and ε_i^* the residual. It is assumed that η^* and ε_i^* are independent, and that $\varepsilon_i^* \sim N(0, \sigma_{\varepsilon^*}^2)$ and $\eta^* \sim SN(\kappa, \omega, \zeta)$, with location parameter κ , scale parameter ω , and shape parameter ζ (Azzalini, 1985, 1986). The probability density function of the skew-normal factor scores η^* is the following:

$$f(x|\kappa, \omega, \zeta) = \frac{2}{\omega} \Phi\left(\zeta \frac{x - \kappa}{\omega}\right) \phi\left(\frac{x - \kappa}{\omega}\right), \quad (2)$$

where ω is the scale parameter, $\Phi(\cdot)$ the standard normal distribution function, ζ the shape parameter, κ the location parameter and $\phi(\cdot)$ the standard normal probability density

function. Note that the normal distribution is a special case of the skew-normal distribution, with $\zeta = 0$.

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

$$f(\mathbf{t}) = 2\phi_Q(\mathbf{t}; \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}'\mathbf{t}), \quad (3)$$

where $\phi_Q(\mathbf{t}; \boldsymbol{\Omega})$ is the Q -dimensional normal density with mean zero and correlation matrix $\boldsymbol{\Omega}$, and where the vector $\boldsymbol{\alpha}$ contains the shape parameters (related to ζ in equation (2)).

The well-known equivalence of the normal factor model 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 and Capitanio (1999) show that for each $\boldsymbol{\eta}^*$ 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 $\boldsymbol{\alpha}' = (\alpha_1 \ 0 \ \dots \ 0)$. Thus, the first random variable is then a one-dimensional 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 equation (3) 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 the factors are orthogonal, and only a single factor has a shape parameter $\alpha \neq 0$ (Montanari & Viroli, 2010). The model can be fitted by means of marginal maximum likelihood (Bock & Aitkin, 1981; Molenaar *et al.*, 2010; Montanari & Viroli, 2010).

2.2. The quadratic factor model

The quadratic factor model is an alternative approach to account for non-normally distributed conditional means (Molenaar *et al.*, 2010). It is a special case of the non-linear factor model (McDonald, 1962, 1967; Mooijaart & Bentler, 1986), which we will first describe.

If y_i denotes a randomly observed score on item i , the following non-linear factor model (McDonald, 1962, 1967; Mooijaart & Bentler, 1986) is specified for y_i :

$$y_i = \tilde{v}_i + \tilde{\lambda}_i s(\eta) + \varepsilon_i, \quad (4)$$

where \tilde{v}_i is the intercept of item i , $\tilde{\lambda}_i$ the factor loading, η the common factor score, $s(\eta)$ a function of the factor scores and ε_i the residual. It is assumed that $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, $\eta \sim N(\mu, \sigma_\eta^2)$, and that η and ε_i are independent. For $s(\eta)$ one may specify a polynomial function

$$s(\eta) = \gamma_{i0} + \gamma_{i1}\eta + \gamma_{i2}\eta^2 + \dots + \gamma_{ir}\eta^r. \quad (5)$$

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

$$y_i = v_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2 + \varepsilon_i, \quad (6)$$

where v_i is the intercept, 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 methods based on maximum likelihood estimation (Harring, Weiss, & Hsu, 2012; Klein & Muthén, 2007; Rizopoulos & Moustaki, 2008).

2.3. Relationship between the skew-normal factor model and the quadratic factor model

We will now show that the distribution of the items under the quadratic factor model and the skew-normal factor model is equivalent up to third-order moments, but that the converse is not generally true.

We consider the skew-normal factor model in equation (1) and the quadratic factor model in equation (6). 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 of 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 conditional means of the two models.

Under the skew-normal factor model T_i^* , the conditional mean $E(y_i | \eta^*)$, equals

$$T_i^* = v_i^* + \lambda_i^* \eta^*, \quad (7)$$

where $\eta^* \sim SN(\kappa, \omega, \zeta)$, with location parameter κ , scale parameter ω , and shape parameter ζ . To identify the model, and without loss of generality, we fix κ at 0 and ω at 1, so that $T_i^* = v_i^* + \lambda_i^* \eta^*$ with $\eta^* \sim SN(0, 1, \zeta)$.

Under the quadratic factor model, T_i , the conditional mean $E(y_i | \eta)$, equals

$$T_i = v_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2, \quad (8)$$

where $\eta \sim N(\mu, \sigma_\eta^2)$. To identify the model, we fix μ at 0 and σ_η^2 at 1, so that $T_i = v_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2$ with $\eta \sim N(0, 1)$. To have fully equivalent models, the densities of T_i and T_i^* should be equal.

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

As we will show, the density of T_i (under the quadratic factor model) can be made equivalent up to its third moment to that of T_i^* (under the skew-normal factor model). For the sake of simplicity, we consider a special case of the skew-normal factor model for T_i^* by fixing v_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 v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$.

The question now reduces to whether there exist constants v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$ which equate the first three moments of the density of $T_i = v_i + \lambda_{i(1)}\eta + \lambda_{i(2)}\eta^2$ (equation 8) to the density of η^* , with $\eta^* \sim SN(\kappa, \omega, \zeta)$. In Appendix A, it is proved that those constants exist (omitting the index i in v_i , $\lambda_{i(1)}$, $\lambda_{i(2)}$ and T_i to improve readability). To find the constants v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$, one first needs to find $\lambda_{i(2)}$ (by solving equation [A6] in Appendix A for λ_2 satisfying $1 - c_1^2 - 2\lambda_{(2)}^2 > 0$). Then $\lambda_{i(1)}$ and v_i can be computed as

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

Further, it is shown in Appendix A that $\lambda_{i(2)}$ and v_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 foregoing implies that the density of T_i^* (under the skew-normal factor model) can be approximated closely by that of T_i (under the quadratic factor model). To illustrate how well this approximation is, we consider three cases with $\zeta = 1.81$, 2.17 and 3.50, corresponding to a small, medium and large coefficient, respectively (see, e.g., Molenaar *et al.*, 2010, for an indication of the magnitude of shape parameters). To assess the closeness of the densities of T_i^* and T_i , we use the L_1 -norm of the difference between their densities:

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

The L_1 -norm of the density differences and the values for v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$ can be found in Table 1. As can be seen, the L_1 -norm is rather small, even for large values of ζ . The closeness of the three distributions of T_i^* and T_i is further illustrated in Figure 1(a–c). As can be seen, they are very close.

To further illustrate the closeness of the distributions, Figure 1(d) illustrates the L_1 -norm of the density differences for values of ζ between -3 and 3 . As can be seen, the L_1 -norm is very small in the range of parameter values of practical interest. 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 why, as demonstrated by Molenaar *et al.* (2010), a quadratic factor model cannot include a skew-normal factor, because such a model would not be empirically identified.

One may ask how well T_i^* could be approximated by a non-linear factor model with a polynomial function of a degree larger than two. We conjecture that increasing the degree of the polynomial will improve the approximation. 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 $k = 1, 2, \dots$, then the distribution of T_i will converge to the distribution of η^* (Billingsley, 1995, Section 30). Therefore, taking the degree of the polynomial T_i to be of a degree 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 T_i^* and T_i will be even better than we already had with a second-degree polynomial.

Table 1. The L_1 -norm of the density differences of T_i^* and T_i , and the values for v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$, for different values of ζ

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

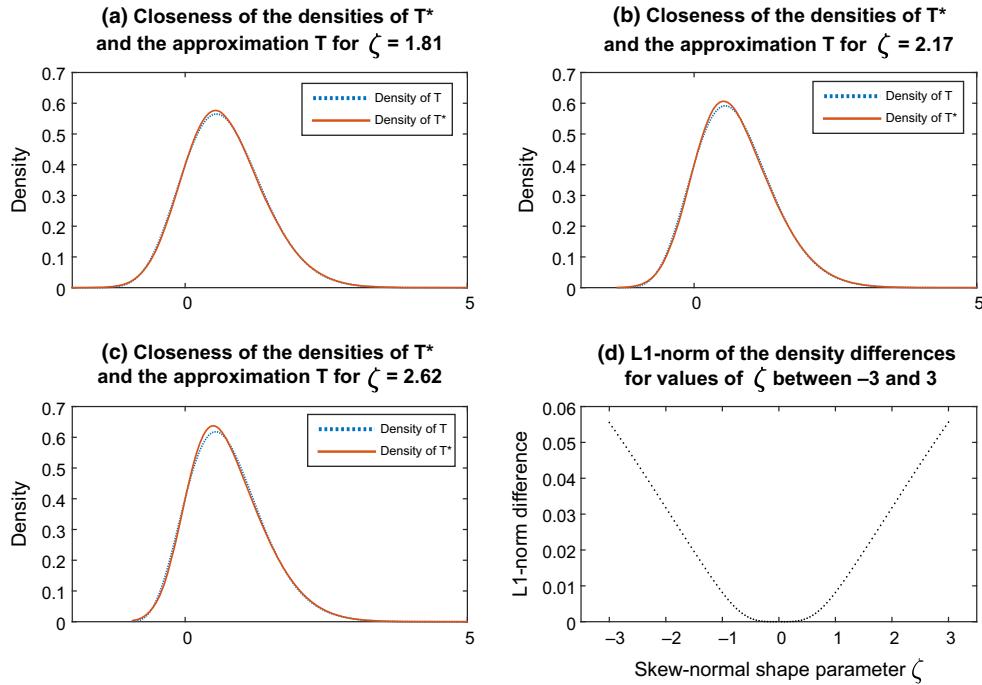


Figure 1. Closeness of the distributions of T_i^* and the approximation T_i for (a) a small ($\zeta = 1.81$), (b) medium ($\zeta = 2.17$), and (c) large ($\zeta = 2.62$) skewness coefficient, and (d) the L_1 -norm of these density differences for values of ζ between -3 and 3 .

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

We first consider whether, for a single item i that follows a quadratic factor model, there exists a skew-normal factor model representation, in terms of equality of the first three moments of their densities. In Appendix B, it is proved that this cannot be done in all cases. There are two limiting factors.

First, equating the densities appears to be limited by the range of skewness (and kurtosis) of the skew-normal distribution (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 falls 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 . Second, the skew-normal factor model representation only exists if the quadratic terms are equal to each other for all items.

To further illustrate how well the density of T_i can be approximated by that of T_i^* , if $\lambda_{i(2)}$ is small and equal across all items, we assess the closeness of the densities using the L_1 -norm of their difference. Figure 2 illustrates the L_1 -norm of the density differences for values of $\lambda_{i(2)}$ between -0.15 and 0.15 . As can be seen, the L_1 -norm is very small in this range. This implies that, for those small and equal quadratic terms, the two models would be empirically indistinguishable from each other, for the parameters of practical interest.

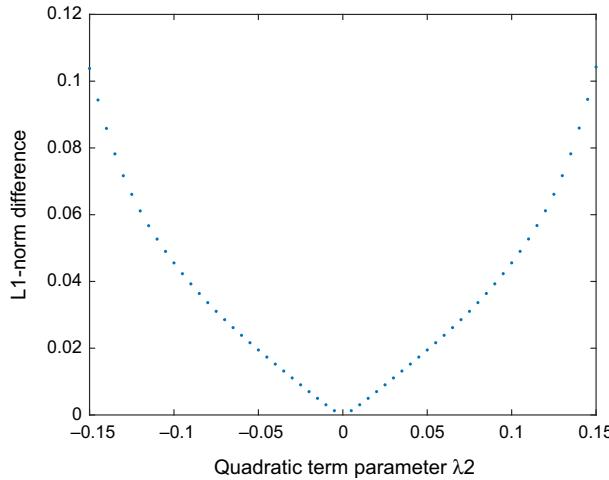


Figure 2. L_1 -norm of the density differences for values of $\lambda_{(2)}$ between -0.15 and 0.15 .

2.3.3. 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 quadratic loadings are equal across items loading on that factor and within certain bounds. In Appendix D, it is shown that this relationship between the skew-normal factor model and the non-linear factor model for a single factor can be generalized to the case with multiple factors. That is, if the conditional mean in the skew-normal factor model is

$$T_i^* = v_i^* + \lambda_{i(1)}^* \eta_1^* + \sum_{q=2}^Q \lambda_{i(q)}^* \eta_q^*,$$

where the Q -dimensional skew-normal distribution is in canonical form (with η_1^* skew-normal and $\eta_q^* \sim N(0, 1)$, $q = 2, \dots, Q$), then we can find parameters v_i , $\lambda_{i(1)}$ and $\lambda_{i(2)}$, $\lambda_{i(q)}$, $q = 2, \dots, Q$, such that the non-linear factor model with conditional mean $T_i = v_i + \lambda_{i(1)} \eta_1 + \lambda_{i(2)} \eta_1^2 + \sum_{q=2}^Q \lambda_{i(q)} \eta_q$ (with η_q , $q = 1, \dots, Q$, independent $N(0, 1)$ variables) satisfies $E(T_i^{*k}) = E(T_i^k)$, for $k = 1, 2, 3$.

This generalization of the relationship between the non-linear and skew-normal factor models 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 non-linear factor model, and conversely if the skewness of T_i does not fall outside the range of the skewness of T_i^* and is equal across all items.

2.4. Implications of the relationship between the skew-normal 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 quadratic loadings are small and equal across items loading on that factor. From a mathematical point of view, in the conditions mentioned,

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, yielding more efficient estimates. Moreover, one uses linear relations between the items and the latent trait, which are generally easier to interpret than non-linear relations. Furthermore, the linear factor model is asymptotically robust to non-normality (Amemiya & Anderson, 1990; Anderson & Amemiya, 1988). The quadratic factor model is more flexible than the skew-normal factor model. For example, it allows for different degrees of skewness of the different items, unlike the skew-normal factor model. Therefore, the quadratic factor model is capable of describing a wider range of data.

3. Empirical example

To give an example of how one may choose between the different models in empirical practice, we present an application for data from the clinical screening instrument Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1977, 1994). We analysed data from a group of psychiatric outpatients who completed the SCL-90-R during admission to a mental health clinic or university research clinic in the Netherlands. The sample analysed consists of $N = 1,842$ psychiatric outpatients with a mean age of 35.2 ($SD = 11.0$), consisting of 729 males, 1,109 females and four of unknown gender. Further details on the sample (including exclusion criteria) can be found in Smits, Timmerman, Barelds, and Meijer (2014). For illustrative purposes, we considered the eight subscale scores of the SCL-90-R: Agoraphobia, Anxiety, Depression, Somatization, Cognitive Performance Deficits, Interpersonal Sensitivity, Hostility, and Sleep Difficulties. The analyses were conducted in Mx (see Molenaar *et al.*, 2010, for details on such analyses and example scripts).

We fitted a linear factor model, a quadratic factor model and a skew-normal factor model to these data. The fits of the models are presented in Table 2. We considered the likelihood ratio test (LRT) to compare the fit of the baseline model (the linear factor model) to those of the quadratic factor model and the skew-normal factor model. The latter two were compared using the Akaike information criterion (AIC), Bayesian information criterion (BIC), sample size adjusted BIC (SABIC) and deviance information criterion (DIC) fit indices.

As can be seen in Table 2, accounting for non-normality of the data by allowing a non-linear factor to item relationship improved the fit significantly, $\chi^2(8) = 339.83, p < .005$.

Table 2. Fit indices for the three models fitted

Factor model	-2 LL	df	LRT	Δ df	p-Value	AIC	BIC	SABIC	DIC
Linear (baseline model)	95,751.29	14,712				66,327.29	-7,431.23	15,938.55	6,088.20
Quadratic	95,411.46	14,704	339.83	8	<.005	66,003.46	-7,571.07	15,786.00	5,941.00
Skew-normal	95,616.64	14,711	134.65	1	<.005	66,194.64	-7,494.79	15,873.40	6,023.71

Notes. $-2 LL = -2$ times the log-likelihood; LRT = likelihood ratio test statistic between that model and the baseline model; AIC = Akaike information criterion; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; DIC = deviance information criterion.

This suggests that the assumption of normally distributed subscale scores is violated. If this non-linearity is not substantial and about equal across subscales, one may as well take account for this non-linearity by allowing for a skewed distributed factor. As can be seen in Table 2, the skew-normal factor model fits the data significantly better than the linear factor model, $\chi^2(1) = 134.65, p < .005$. However, according to the AIC, BIC, SABIC and DIC, the quadratic factor model is preferred over the skew-normal factor model, implying that non-linear scale to factor relationships are needed to describe the non-normality in the data. In Table 3, the parameter estimates of the models are presented. As can be seen, the quadratic factor loadings $\lambda_{i(2)}$ are relatively large and vary reasonably across subscales, illustrating the need to model the non-linearity in the data through a quadratic factor model instead of a skew-normal factor model.

To conclude, the results suggest that for the subscale scores of the SCL-90-R, the assumption of normally distributed scale scores seems untenable but should be accounted for. Furthermore, the comparative fit measures suggest that this non-normality can be best accounted for by allowing non-linear factor to scale relationships, since the non-linearity in the data seems to differ too much across subscale scores.

4. Discussion

Deviations from normality of the conditional means can be modelled through either a skew-normal factor model or a quadratic factor model (Molenaar *et al.*, 2010). In this paper we showed why these two variants to account for non-normal conditional means cannot be implemented jointly in a single model. We showed and illustrated that the quadratic factor model is equivalent to the skew-normal factor model up to third-order moments, and that the converse is only true if the factor loading of the quadratic term is small and equal across items. Furthermore, the intimate relationship between the skew-normal factor model and the quadratic factor model holds for both the single and multiple factor case. This has the following implications for their use in empirical practice. Observed data that follow any skew-normal factor model can be so well approximated

Table 3. Parameter estimates for the three models fitted

Factor model	Ag	An	De	So	Co	In	Ho	Sl
Linear								
$\lambda_{i(1)}$	4.32	7.63	12.02	7.71	5.98	11.61	3.27	2.21
v_i	12.92	23.89	43.25	25.75	22.23	40.13	11.94	7.84
σ_e^2	22.95	24.24	57.51	39.01	22.35	81.95	18.81	9.62
Quadratic								
$\lambda_{i(1)}$	4.14	7.47	12.06	7.50	5.94	11.36	3.16	2.21
$\lambda_{i(2)}$	1.40	1.12	0.25	1.45	0.35	1.65	0.69	0.11
v_i	11.52	22.77	43.00	24.29	21.88	38.48	11.24	7.74
σ_e^2	20.92	23.92	56.18	37.88	22.46	82.11	18.60	9.62
Skew-normal								
ζ	4.39							
$\lambda_{i(1)}^*$	4.45	7.75	12.02	7.85	6.00	11.73	3.32	2.23
v_i^*	12.87	23.81	43.13	25.67	22.17	40.01	11.90	7.82
σ_e^*	22.28	23.66	60.82	38.14	22.86	82.12	18.74	9.67

Notes. Ag = Agoraphobia; An = Anxiety; De = Depression; So = Somatization; Co = Cognitive Performance Deficits; In = Interpersonal Sensitivity; Ho = Hostility; Sl = Sleep Difficulties.

with the quadratic factor model that the models are indistinguishable in practice. Further, observed data that follow a quadratic factor model, with quadratic terms that are small and similar across items, can be very well approximated with a skew-normal factor model. In those conditions where the two models are indistinguishable, the skew-normal factor model is generally preferred for reasons of parsimony and interpretability of the linear versus non-linear relationships.

The quadratic factor model is more flexible than the skew-normal factor model. If the values for the quadratic term are too large or differ too much from item to item, then one is bound to model this non-normality with a quadratic factor model. We recommend researchers to use a skew-normal factor model in conditions where the two models are empirically indistinguishable and to move to a quadratic factor model when a more flexible model is needed.

When observed data comply better with a skew-normal than with a normal factor model we favour the use of the skew-normal factor model for reasons of interpretability. Further, though the maximum likelihood estimation of the linear factor model is asymptotically robust to non-normality (Amemiya & Anderson, 1990; Anderson & Amemiya, 1988) and yields consistent estimates for data following a skew-normal factor model (Shapiro, 1984), its behaviour with limited sample sizes may be problematic. When data follow the corresponding quadratic factor model, parameter estimates of the linear factor model are biased (Bauer, 2005), ruling out the latter option in those cases.

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. This could be accounted for by including higher-order polynomials in a non-linear factor model. Further, we note that an extension of the skew-normal distribution exists in which an additional shape parameter is included such that the range of the skewness of the distribution is wider (Azzalini, 1985; Henze, 1986). We expect that for such an extended skew-normal factor model the first three moments can be equated to the first three moments of the quadratic factor model for a wider range of factor loadings of the quadratic term. However, such an extended skew-normal factor model will still be less flexible than the quadratic factor model since the latter features varying quadratic term loadings from item to item.

Acknowledgements

The research of Alwin Stegeman is supported by the Dutch Organization for Scientific Research (NWO), VIDI grant 452-08-001. We would like to thank W.A. Arrindell for sharing the data of the SCL-90.

References

- Amemiya, Y., & Anderson, T. W. (1990). Asymptotic chi-square tests for a large class of factor-analysis models. *Annals of Statistics*, 18, 1453–1463. doi:10.1214/aos/1176347760
- Anderson, T. W., & Amemiya, Y. (1988). The asymptotic normal-distribution of estimators in factor-analysis under general conditions. *Annals of Statistics*, 16, 759–771. doi:10.1214/aos/1176350834
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171–178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32, 159–188. doi:10.1111/j.1467-9469.2005.00426.x

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, Series B*, 61, 579–602. doi:10.1111/1467-9868.00194

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726. doi:10.1093/biomet/83.4.715

Bauer, D. J. (2005). The role of nonlinear factor-to-indicator relationships in tests of measurement equivalence. *Psychological Methods*, 10, 305–316. doi:10.1037/1082-989X.10.3.305

Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York, NY: Wiley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459. doi:10.1007/BF02293801

Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring & procedures manual-II for the R(evised) version*. Towson, MD: Clinical Psychometric Research.

Derogatis, L. R. (1994). *SCL90-R: Administration, scoring and procedures manual*. Minneapolis, MN: National Computer Systems.

Gupta, A. K., Nguyen, T. T., & Sanqui, J. A. T. (2004). Characterization of the skew-normal distribution. *Annals of the Institute of Statistical Mathematics*, 56, 351–360. doi:10.1007/BF02530549

Harring, J. R., Weiss, B. A., & Hsu, J. (2012). A comparison of methods for estimating quadratic effects in nonlinear structural equation models. *Psychological Methods*, 17, 193–214. doi:10.1037/a0027539

Henze, N. (1986). A probabilistic representation of the ‘skew-normal’ distribution. *Scandinavian Journal of Statistics*, 13, 271–275.

Hessen, D. J., & Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology*, 62, 57–77. doi:10.1348/000711007X248884

Irving, R. S. (2004). *Integers, polynomials, and rings*. New York, NY: Springer-Verlag.

Jedidi, K., Jagpal, H., & DeSarbo, W. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39–59. doi:10.1287/mksc.16.1.39

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457–474. doi:10.1007/BF02296338

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, 42, 647–673. doi:10.1080/00273170701710205

Ma, Y., & Genton, M. G. (2010). Explicit estimating equations for semiparametric generalized linear latent variable models. *Journal of the Royal Statistical Society, Series B*, 72, 475–495. doi:10.1111/j.1467-9868.2010.00741.x

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika*, 27, 392–415. doi:10.1007/BF02289646

McDonald, R. P. (1967). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, 32, 77–112. doi:10.1007/BF02289406

Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, 63, 293–317. doi:10.1348/000711009X456935

Montanari, A., & Viroli, C. (2010). A skew-normal factor model for the analysis of student satisfaction towards university courses. *Journal of Applied Statistics*, 37, 473–487. doi:10.1080/02664760902736737

Mooijaart, A. (1985). Factor analysis for non-normal variables. *Psychometrika*, 50, 323–342. doi:10.1007/BF02294108

Mooijaart, A., & Bentler, P. (1986). Random polynomial factor analysis. In E. Diday, Y. Escoufier, L. Lebart, J. Pages & Y. Schekhtman *et al.* (Eds.), *Data analysis and informatics, IV* (pp. 241–250). Amsterdam, the Netherlands: Elsevier.

Rizopoulos, D., & Moustaki, I. (2008). Generalized latent variable models with non-linear effects. *British Journal of Mathematical and Statistical Psychology*, 61, 415–438. doi:10.1348/000711007X213963

Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical & Statistical Psychology*, 37, 84–88. doi:10.1111/j.2044-8317.1984.tb00790.x

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Smits, I. A. M., Timmerman, M. E., Barelds, D. P. H., & Meijer, R. R. (2015). The Dutch Symptom Checklist-90-Revised: Is the use of the subscales justified? *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000233

Received 10 September 2013; revised version received 14 September 2015

Appendix A

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

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

Hence, we obtain

$$v = c_1 - \lambda_{(2)}. \quad (\text{A1})$$

For $k = 2$, we have

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

Hence, we obtain

$$v^2 + 2v\lambda_{(2)} + 3\lambda_{(2)}^2 + \lambda_{(1)}^2 - 1 = 0. \quad (\text{A2})$$

For $k = 3$, we have

$$\begin{aligned} E(\eta^{*3}) &= \sqrt{\frac{2}{\pi}} \frac{3\zeta}{(1 + \zeta^2)^{3/2}} \left(1 + \frac{4\zeta^2}{6} \right) = c_3, \\ E(T^3) &= v^3 + 3v^2\lambda_{(2)} + 3v\lambda_{(1)}^2 + 9v\lambda_{(2)}^2 + 9\lambda_{(1)}^2\lambda_{(2)} + 15\lambda_{(2)}^3. \end{aligned}$$

Hence, we obtain

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

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

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

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

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 into (A5). After simplifying, this yields

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

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 v can be computed as

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

Here, both $\lambda_{(2)}$ and v 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., Irving, 2004, Section 10.3).

For the polynomial in (A6), the discriminant depends on ζ . We have

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

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

$$3,456\pi^3 + 10,368(-2\pi^2 + \pi^3)\zeta^2 + 10,368(4\pi - 4\pi^2 + \pi^3)\zeta^4 + 864(-48 + 56\pi - 25\pi^2 + 4\pi^3)\zeta^6.$$

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 $D(0) = 3,456$, we have proved 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

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

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

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

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. \square

Appendix B

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

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

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

For the left-hand side, we compute

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

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

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

As $|\lambda_{(2)}|$ 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 $|\lambda_{(2)}|$ 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.

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 $v = 0$ and $\lambda_{(1)} = 1$ without loss of generality, since the location and scaling can be absorbed in the parameters v^* 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*

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

Let T^ have the skew-normal distribution with parameters (v^*, λ^*, ζ) . Then there exist parameters v^*, 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

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

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

In Appendix B, 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

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

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 v^* and λ^* . Since the skewnesses of T and T^* are equal, it then also follows that $E(T^3) = E(T^{*3})$.

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

$$v^* = \lambda_{(2)} - \sqrt{\frac{2}{\pi}}\delta\lambda^*. \quad (\text{C5})$$

Setting $E(T^2) = E(T^{*2})$ in (C3) and substituting (C5) for v^* 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

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

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

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 η^* have a Q -dimensional skew-normal distribution, defined by (5), in the canonical form with $\Omega = I_Q$ and $\alpha' = (\alpha_1 \ 0 \ \dots \ 0)$, with mean vector and covariance matrix (Montanari & Viroli, 2010)*

$$\mu_{\eta^*} = E(\eta^*) = \sqrt{\frac{2}{\pi}}\delta, \quad \text{Var}(\eta^*) = \Omega - \mu_{\eta^*}\mu'_{\eta^*},$$

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

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

Let

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

where η_1, \dots, η_Q are mutually independent $N(0, 1)$ variables. Then, for any $\alpha_1, v^*, \lambda_1^*, \dots, \lambda_Q^*$ there exist constants $v, \lambda_{11}, \lambda_{12}, \lambda_2, \dots, \lambda_Q$ such that $E(T^{*k}) = E(T^k)$ for $k = 1, 2, 3$.

Proof. Without loss of generality we set $v^* = 0$ and $\lambda_1^* = 1$. Note that η_1^* has a one-dimensional skew-normal distribution with parameters $(0, 1, \alpha_1)$, and $\eta_q^*, q = 2, \dots, Q$, are $N(0, 1)$ distributed. Moreover, $\eta_1^*, \dots, \eta_Q^*$ are mutually independent. Let $T_0^* = \eta_1^*$ and $T_0 = v + \lambda_{11}\eta_1 + \lambda_{12}\eta_1^2$. From Proposition 1 we know that for any α_1 there exist $v, \lambda_{11}, \lambda_{12}$, such that $E(T_0^{*k}) = E(T_0^k)$ for $k = 1, 2, 3$. Let $v, \lambda_{11}, \lambda_{12}$ 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 $q = 2, \dots, Q$. 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. \square

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.