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.

1. 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...
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 Virolé (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:

\[
y_i = \eta^*_i + \lambda^*_i \eta^* + \varepsilon^*_i,
\]

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

\[
f(x|\kappa, \omega, \zeta) = \frac{2}{\omega} \Phi\left( \frac{x - \kappa}{\omega} \right) \varphi\left( \frac{x - \kappa}{\omega} \right),
\]

where \( \omega \) is the scale parameter, \( \Phi(\cdot) \) the standard normal distribution function, \( \zeta \) the shape parameter, \( \kappa \) the location parameter and \( \varphi(\cdot) \) 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 \( (Q, \text{with } Q \geq 1; \text{Montanari & Viroli, 2010}) \), which involves the multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996). Specifically, it involves the \( (Q \times 1) \) vector with factor scores \( \eta^* \sim SN(\Omega, \alpha) \), with the density function

\[
f(t) = 2\phi_Q(t; \Omega)\Phi(\alpha' t),
\]

where \( \phi_Q(t; \Omega) \) is the \( Q \)-dimensional normal density with mean zero and correlation matrix \( \Omega \), and where the vector \( \alpha \) contains the shape parameters (related to \( \zeta \) 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 \( \eta^* \sim SN(\Omega, \alpha) \), and \( H \) is a non-singular \( (Q \times Q) \) matrix such that \( H'\Omega H \) is a correlation matrix, then \( H'\eta^* \sim SN(H'\Omega H, H'\Omega H^{-1}) \). Azzalini and Capitanio (1999) show that for each \( \eta^* \) a linear transformation \( H' \eta^* \) exists that transforms the multivariate skew-normal density to a canonical form with \( \Omega = I_Q \) and \( \alpha' = (\alpha_1, 0 \ldots 0) \). 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 \( y_i \):

\[
y_i = \tilde{\nu}_i + \tilde{\lambda}_i s(\eta) + \varepsilon_i,
\]

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

\[
s(\eta) = \gamma_{i0} + \gamma_{i1} \eta + \gamma_{i2} \eta^2 + \ldots + \gamma_{ir} \eta^r.
\]

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

...
where $v_i$ is the intercept, and $\lambda_{i(k)}$ the factor loading associated with the $k$th power of $\eta$ 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 $\varepsilon_i^*$ and $\varepsilon_i$ rely on exactly the same assumptions in both models, and that the distributions of $\varepsilon_i^*$ and $\varepsilon_i$ 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

$$T_i^* = v_i^* + \lambda_i^* \eta^*,$$

where $\eta^* \sim SN(\kappa, \omega, \zeta)$, with location parameter $\kappa$, scale parameter $\omega$, and shape parameter $\zeta$.

To identify the model, and without loss of generality, we fix $\kappa$ at 0 and $\omega$ 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

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

where $\eta \sim N(\mu, \sigma_\eta^2)$. To identify the model, we fix $\mu$ at 0 and $\sigma_\eta^2$ 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 $\lambda_i^*$ 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 $\eta^*$, 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 $\lambda_{i(2)}$ satisfying $1 - c_1^2 - 2c_2^2 > 0$). Then $\lambda_{i(1)}$ and $v_i$ can be computed as
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:

$$\|f_{T^{*}}(y) - f_{T}(y)\|_{1} = \int |f_{T^{*}}(y) - f_{T}(y)| dy.$$ 

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 $\zeta$. 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 $\zeta$ 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 $E(T_{i}^{k})$ converge to the moments of $E(\eta^{k})$ for $k = 1, 2, \ldots$, then the distribution of $T_{i}$ will converge to the distribution of $\eta^{*}$ (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 $E(T_{i}^{k}) = E(\eta^{k})$, 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>
<thead>
<tr>
<th>$\zeta$</th>
<th>$L_1$-norm</th>
<th>$v_{i}$</th>
<th>$\lambda_{i(1)}$</th>
<th>$\lambda_{i(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.81</td>
<td>0.0271</td>
<td>0.6507</td>
<td>0.7125</td>
<td>0.0477</td>
</tr>
<tr>
<td>2.17</td>
<td>0.0360</td>
<td>0.6671</td>
<td>0.6843</td>
<td>0.0576</td>
</tr>
<tr>
<td>2.62</td>
<td>0.0468</td>
<td>0.6783</td>
<td>0.6598</td>
<td>0.0671</td>
</tr>
</tbody>
</table>
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 $k_i(2)$ the skewness of $T_i$ is outside the range of the skewness of $T_i/C_3$:

Consequently, for these values of $k_i(2)$, the first three moments of the density of $T_i/C_3$ cannot be made equal to those of the density of $T_i$. In Appendix C it is shown that for small values of $k_i(2)$ (roughly between $0.17$ and $0.17$), for which the skewness of $T_i$ falls within the range of the skewness of $T_i/C_3$, the first three moments of $T_i/C_3$ 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 these density differences for values of $\zeta$ 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 these 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 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 $\zeta$ between $-3$ and $3$.
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

\[ T_i = m_i + k_i \alpha^1 \eta^1_i + \sum_{q=2}^Q k_i \alpha^q \eta^q_i, \]

where the \( Q \)-dimensional skew-normal distribution is in canonical form (with \( \eta^1_i \) skew-normal and \( \eta^q_i \sim N(0, 1), q = 2, \ldots, Q \), then we can find parameters \( v_i, \lambda_{i(1)} \) and \( \lambda_{i(2)} \), \( \lambda_{i(q)}, q = 2, \ldots, Q \), such that the non-linear factor model with conditional mean \( T_i = v_i + \lambda_{i(1)} \eta^1_i + \lambda_{i(2)} \eta^2_i + \sum_{q=2}^Q \lambda_{i(q)} \eta^q_i \) (with \( \eta^q_i, q = 1, \ldots, Q \), 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 \( T_i^k \) 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>
<thead>
<tr>
<th>Factor model</th>
<th>(-2 LL)</th>
<th>df</th>
<th>LRT</th>
<th>(\Delta df)</th>
<th>(p)-Value</th>
<th>AIC</th>
<th>BIC</th>
<th>SABIC</th>
<th>DIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear (baseline model)</td>
<td>95,751.29</td>
<td>14,712</td>
<td></td>
<td></td>
<td></td>
<td>66,327.29</td>
<td>-7,431.23</td>
<td>15,938.55</td>
<td>6,088.20</td>
</tr>
<tr>
<td>Quadratic</td>
<td>95,411.46</td>
<td>14,704</td>
<td>339.83</td>
<td>8</td>
<td>&lt;.005</td>
<td>66,003.46</td>
<td>-7,571.07</td>
<td>15,786.00</td>
<td>5,941.00</td>
</tr>
<tr>
<td>Skew-normal</td>
<td>95,616.64</td>
<td>14,711</td>
<td>134.65</td>
<td>1</td>
<td>&lt;.005</td>
<td>66,194.64</td>
<td>-7,494.79</td>
<td>15,873.40</td>
<td>6,023.71</td>
</tr>
</tbody>
</table>

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 $\lambda_{k(2)}$ 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 (Molenar 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

<table>
<thead>
<tr>
<th>Factor model</th>
<th>Ag</th>
<th>An</th>
<th>De</th>
<th>So</th>
<th>Co</th>
<th>In</th>
<th>Ho</th>
<th>Sl</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{r(1)}$</td>
<td>4.32</td>
<td>7.63</td>
<td>12.02</td>
<td>7.71</td>
<td>5.98</td>
<td>11.61</td>
<td>3.27</td>
<td>2.21</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>12.92</td>
<td>23.89</td>
<td>43.25</td>
<td>25.75</td>
<td>22.23</td>
<td>40.13</td>
<td>11.94</td>
<td>7.84</td>
</tr>
<tr>
<td>$\sigma^2_\xi$</td>
<td>22.95</td>
<td>24.24</td>
<td>57.51</td>
<td>39.01</td>
<td>22.35</td>
<td>81.95</td>
<td>18.81</td>
<td>9.62</td>
</tr>
<tr>
<td><strong>Quadratic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{r(1)}$</td>
<td>4.14</td>
<td>7.47</td>
<td>12.06</td>
<td>7.50</td>
<td>5.94</td>
<td>11.36</td>
<td>3.16</td>
<td>2.21</td>
</tr>
<tr>
<td>$\lambda_{r(2)}$</td>
<td>1.40</td>
<td>1.12</td>
<td>0.25</td>
<td>1.45</td>
<td>0.35</td>
<td>1.65</td>
<td>0.69</td>
<td>0.11</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>11.52</td>
<td>22.77</td>
<td>43.00</td>
<td>24.29</td>
<td>21.88</td>
<td>38.48</td>
<td>11.24</td>
<td>7.74</td>
</tr>
<tr>
<td>$\sigma^2_\xi$</td>
<td>20.92</td>
<td>23.92</td>
<td>56.18</td>
<td>37.88</td>
<td>22.46</td>
<td>82.11</td>
<td>18.60</td>
<td>9.62</td>
</tr>
<tr>
<td><strong>Skew-normal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\zeta$</td>
<td>4.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{r(1)}$</td>
<td>4.45</td>
<td>7.75</td>
<td>12.02</td>
<td>7.85</td>
<td>6.00</td>
<td>11.73</td>
<td>3.32</td>
<td>2.23</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>12.87</td>
<td>23.81</td>
<td>43.13</td>
<td>25.67</td>
<td>22.17</td>
<td>40.01</td>
<td>11.90</td>
<td>7.82</td>
</tr>
<tr>
<td>$\sigma^2_\xi$</td>
<td>22.28</td>
<td>23.66</td>
<td>60.82</td>
<td>38.14</td>
<td>22.86</td>
<td>82.12</td>
<td>18.74</td>
<td>9.67</td>
</tr>
</tbody>
</table>

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


Appendix A

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

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

Hence, we obtain

\[ \nu = c_1 - \lambda_{(2)}. \quad (A1) \]

For \( k = 2 \), we have

\[ E(\eta^{*2}) = 1, \quad E(T^2) = \nu^2 + 2\nu\lambda_{(2)} + 3\lambda_{(2)}^2 + \lambda_{(1)}^2. \]

Hence, we obtain

\[ \nu^2 + 2\nu\lambda_{(2)} + 3\lambda_{(2)}^2 + \lambda_{(1)}^2 - 1 = 0. \quad (A2) \]

For \( k = 3 \), we have

\[ E(\eta^{*3}) = \sqrt{\frac{2}{\pi}} \frac{3\zeta}{(1 + \zeta^2)^{3/2}} \left( 1 + \frac{4\zeta^2}{6} \right) = c_3, \]

\[ E(T^3) = \nu^3 + 3\nu^2\lambda_{(2)} + 3\nu\lambda_{(1)}^2 + 9\nu\lambda_{(2)}^2 + 9\lambda_{(1)}\lambda_{(2)} + 15\lambda_{(2)}^3. \]

Hence, we obtain
\[ v^3 + 3v^2 \lambda_{(2)} + 3v \lambda_{(1)}^2 + 9v \lambda_{(1)}^2 \lambda_{(2)} + 9 \lambda_{(1)}^2 \lambda_{(2)}^2 + 15 \lambda_{(2)}^3 - c_3 = 0. \]  \hspace{1cm} (A3)

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

\[
2 \lambda_{(2)}^2 + \lambda_{(1)}^2 + c_1^2 - 1 = 0, \hspace{1cm} (A4)
\]

\[
8 \lambda_{(2)}^3 + 6c_1 \lambda_{(2)}^2 + 6 \lambda_{(1)} \lambda_{(2)} + 3c_1 \lambda_{(1)}^2 + c_1^3 - c_3 = 0. \hspace{1cm} (A5)
\]

Note that (A4) only has a real solution for \( \lambda_{(1)}, \lambda_{(2)} \) if \( c_1^2 < 1 \). This holds for all \( \zeta \), 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

\[
-4 \lambda_{(2)}^3 + 6(1 - c_1^2) \lambda_{(2)} + 2c_1^4 + 3c_1 - c_3 = 0. \hspace{1cm} (A6)
\]

We show that this third-degree polynomial in \( \lambda_{(2)} \) has three distinct real roots for any shape parameter \( \zeta \), 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

\[
\lambda_{(1)} = \sqrt{1 - c_1^2 - 2 \lambda_{(2)}^2}, \hspace{1cm} v = c_1 - \lambda_{(2)}.
\]

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 \( \lambda_{(1)} v \) 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 \( \zeta \), 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 \( \zeta \). We have

\[
D(\zeta) = 16 \times 6^3(1 - c_1^2)^3 - 27 \times 16(-2c_1^4 + 3c_1 - c_3)^2.
\]

Using symbolic computation software, it can be verified that \( \pi^3(1 + \zeta^2)^3D(\zeta) \) 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 \( \zeta \) has six complex roots. It follows that \( D(\zeta) > 0 \) for any \( \zeta \) if and only if \( D(\zeta) > 0 \) for some \( \zeta \). Since \( D(0) = 3,456 \), we have proved that the polynomial (A6) has three distinct real roots for any \( \zeta \).
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
\[\lambda_{(2)}^{(\text{min})} = -\sqrt{\frac{1 - c_1^2}{2}}, \quad \lambda_{(2)}^{(\text{max})} = \sqrt{\frac{1 - c_1^2}{2}}.\]

Note that \(1 - c_1^2 > 0\) for any \(\zeta\), as shown below (A5). Also note that the coefficient of \(k_2^2\) in (A6) is negative, which implies that \(\lambda_{(2)}^{(\text{min})} < \lambda_{(2)}^{(\text{max})}\).

Since the polynomial (A6) has three real roots, there is exactly one root in between \(\lambda_{(2)}^{(\text{min})}\) and \(\lambda_{(2)}^{(\text{max})}\). We have
\[1 - c_1^2 - 2\left(\lambda_{(2)}^{(\text{min})}\right)^2 = 1 - c_1^2 - 2\left(\lambda_{(2)}^{(\text{max})}\right)^2 = 0.\]

Hence, for the root \(\lambda_{(2)}^*\) in between \(\lambda_{(2)}^{(\text{min})}\) and \(\lambda_{(2)}^{(\text{max})}\) 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 \(\nu, \lambda_{(1)}, \lambda_{(2)}\), there do not exist values for \(\nu^*, \lambda_1^*, \zeta\) that equate the first three moments of \(T^* = \nu^* + \lambda_1^* \zeta^*\) 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 \((\nu^*, \lambda_1^*, \zeta^*)\). 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
\[
E\left(\frac{T - E(T)}{\sqrt{\text{Var}(T)}}\right)^3 = \frac{E(T - E(T))^3}{(E(T - E(T))^2)^{3/2}} = \frac{8\lambda_{(2)}^3 + 6\lambda_{(2)}}{2\lambda_{(2)}^3 + 1}^{3/2}. \quad (B1)
\]

The skewness of \(T^*\) only depends on \(\zeta\). From Azzalini (1985) we obtain
\[
E\left(\frac{T^* - E(T^*)}{\sqrt{\text{Var}(T^*)}}\right)^3 = \left(\frac{4 - \pi}{2}\right)\left(\frac{2}{\pi}\right)^2 \left(\frac{\zeta}{1 + \zeta^2}\right)^3 \left(1 - \frac{2}{\pi}\left(\frac{\zeta^2}{1 + \zeta^2}\right)^{-1}\right)^{-3/2}. \quad (B2)
\]

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

**Proposition 2** Let $T = \eta + \lambda_{(2)} \eta^2$, where $\eta$ has the $N(0, 1)$ distribution and $\lambda_{(2)}$ is a real constant such that
\[ \left| \frac{8\lambda_{(2)}^3 + 6\lambda_{(2)}}{2\lambda_{(2)}^2 + 1} \right| \leq \left( \frac{4 - \pi}{2} \right) \left( \frac{2/\pi}{1 - 2/\pi} \right)^{3/2}. \]  

Let $T^*$ have the skew-normal distribution with parameters $(v^*, \lambda^*, \zeta)$. Then there exist parameters $v^*$, $\lambda^*$, $\zeta$ 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
\[ E(T) = \lambda_{(2)}, \quad E(T^*) = v^* + \lambda^* \sqrt{\frac{2}{\pi}} \frac{\zeta}{1 + \zeta^2}, \]  

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

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 $\zeta$, we estimate $\zeta$ 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 $\gamma$. We substitute $\delta = \zeta / \sqrt{1 + \zeta^2}$. Setting (B2) equal to $\gamma$ and solving for $\delta$ yields
\[ |\delta| = \sqrt{\frac{(\pi/2)|\gamma|^{2/3}}{|\gamma|^{2/3} + (2 - \pi/2)^{2/3}}}, \quad (C4) \]

with \( \delta \) and \( \gamma \) having the same sign. Next, we obtain \( \zeta \) as \( \zeta = \delta/\sqrt{1 - \delta^2} \).

When \( \zeta \) is known, we equate the first and second moments of \( T \) and \( T^* \) to obtain \( v^* \) and \( \lambda^* \). 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

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

Setting \( E(T_2^*) = E(T^{*2}) \) in (C3) and substituting (C5) for \( v^* \) yields, after rewriting,

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

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

\[ \lambda^* = \sqrt{\frac{2\lambda_{(2)}^2 + 1}{1 - 2\delta^2/\pi}}, \quad (C6) \]

Once \( \lambda^* \) 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 \( \eta^* \) have a \( Q \)-dimensional skew-normal distribution, defined by (5), in the canonical form with \( \Omega = I_Q \) and \( \alpha' = (\alpha_1 \ 0 \ldots \ 0) \), 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 \( \eta_1, \ldots, \eta_Q \) are mutually independent \( N(0, 1) \) variables. Then, for any \( \alpha, v^*, \lambda_1^*, \ldots, \lambda_Q^* \) there exist constants \( v, \lambda_{11}, \lambda_{12}, \lambda_2, \ldots, \lambda_Q \) such that \( E(T^{*k}) = E(T^k) \) for \( k = 1, 2, 3 \).

**Proof.** Without loss of generality we set \( v^* = 0 \) and \( \lambda_1^* = 1 \). Note that \( \eta_1^* \) has a one-dimensional skew-normal distribution with parameters \((0, 1, \alpha_1)\), and \( \eta_q^*, q = 2, \ldots, Q \), are \( N(0, 1) \) distributed. Moreover, \( \eta_1^*, \ldots, \eta_Q^* \) 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 \( \alpha_1 \) 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

\[
E(T^{*2}) = E(T_0^{*2} + 2T_0^*(T^* - T_0^*) + (T^* - T_0^*)^2) = E(T_0^{*2}) + 2E(T_0^*)E(T^* - T_0^*) + E(T^* - T_0^*)^2 = E(T_0^2) + \sum_{q=2}^{Q} \lambda_q^{*2},
\]

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 \( q = 2, \ldots, Q \). Then an analogous expansion of \( E(T^{*2}) \) shows that \( E(T^{*2}) = E(T^2) \).

We have

\[
E(T^{*3}) = E(T_0^{*3} + 3T_0^{*2}(T^* - T_0^*) + 3T_0^*(T^* - T_0^*)^2 + (T^* - T_0^*)^3) = E(T_0^{*3}) + 3E(T_0^{*2})E(T^* - T_0^*) + 3E(T_0^*)E(T^* - T_0^*)^2 + E(T^* - T_0^*)^3 = E(T_0^3) + 3E(T_0) \left( \sum_{q=2}^{Q} \lambda_q^{*2} \right),
\]

where we used the independence of \( T_0^* \) and \( T^* - T_0^* \) in the second step, and \( E(T^* - T_0^*) = 0, E(T^{*3}) = E(T_0^{*3}) = E(T_0^3) \) 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.