On the generation of multivariate non-normal distributions by the Fleishman-Vale-Maurelli (FVM) procedure.

This note/letter is meant to clear up my mind about how to investigate the effect of non-normality on the estimation procedures as designed for polynomial factor models, LMS and PLSc\(^1\) in particular. The non-normality I have in mind in this stage pertains to the latent variables; non-normality of the measurement errors in the indicators is less of an issue to me\(^2\). I have a few questions and puzzles and hope you can help me out.

The FVM-procedure (Fleishman (1978) and Vale & Maurelli (1983)) is developed for linear factor models, where all parameters are determined by the covariance matrix of the indicators. They replace the (standard normal) latent variables by ‘well-chosen’ linear combinations of powers of standard normal variables, whose correlations are such that the new latent variables have the same correlations as the original latent variables. ‘Well-chosen’ refers to the satisfaction of the specified requirements concerning the non-normal skewness and (excess-) kurtosis. For \( p \) latent variables\(^3\) \( \eta_k \) this means that we first generate a \( p \)-dimensional normal vector \( z \), with \( z \sim N(0, R_{zz}) \), and then replace \( \eta_k \) by

\[
\tilde{\eta}_k := \sum_{l=1}^{q} w_{k,l} \cdot H_l(z_k) .
\]

The functions \( H_l() \) are ‘orthonormal Hermite polynomials’ (a trivial rewriting and extension of Fleishman’s polynomials). The first 6 orthonormal Hermite polynomials are, with \( x \in \mathbb{R} \),

\[
\begin{align*}
H_1(x) &= x \quad (2) \\
H_2(x) &= (x^2 - 1) \div \sqrt{2} \quad (3) \\
H_3(x) &= (x^3 - 3x) \div \sqrt{6} \quad (4) \\
H_4(x) &= (x^4 - 6x^2 + 3) \div \sqrt{24} \quad (5) \\
H_5(x) &= (x^5 - 10x^3 + 15x) \div \sqrt{120} \quad (6) \\
H_6(x) &= (x^6 - 15x^4 + 45x^2 - 15) \div \sqrt{720} . \quad (7)
\end{align*}
\]

\(^1\)The ‘c’ in PLSc stands for ‘consistent’, meaning that this version of PLS will yield consistent estimators. See Dijkstra (2011) for an outline.

\(^2\)When each (normalized) indicator \( y_{i,j} \) loads on a unique latent variable \( \eta_j \), as in the ‘basic design’, and we keep the normality and the independence (from the latent variable) of the measurement errors, then we get: \( Ey_{i,j}^3 = \lambda_{i,j}^3 \cdot En_i^3 \) and \( Ey_{i,j}^4 - 3 = \lambda_{i,j}^4 \cdot (En_i^4 - 3) \). So the nonnormal skewness and kurtosis do not show themselves as clearly as when we could observe the latent variables directly.

\(^3\)All latent variables will be denoted by \( \eta \) here.
The $H_l(z_k)$'s for $l = 1, 2, \ldots, \infty$ have zero mean, unit variance and they are mutually uncorrelated\(^4\). FVM uses the first three functions to accommodate unit variance and specified levels of skewness($s$) and excess-kurtosis($\kappa$). So the weights $w_{k,l}$ in (1) are in principle determined (though not all values of skewness and kurtosis from the region specified by $\kappa \geq s^2 - 2$ can be attained with $q = 3$). It remains to find the off-diagonal elements of $R_{zz}$, the correlations between the components of $z$. They will have to be such that the covariance matrix of $\tilde{\eta}$ is equal to the covariance matrix of $\eta$. V&M show that the desired/required correlation between $z_i$ and $z_j$ must be the root of a third degree polynomial, with coefficients determined by the weight vectors $w_{i\cdot}$ and $w_{j\cdot}$ and the correlation between $\eta_i$ and $\eta_j$.

The first two questions I have are:

1. Is there always a unique real root of the third degree polynomial, and if not (when we have three real roots), which solution is the right one?
2. Are the solutions always such that $R_{zz}$ is a proper covariance matrix (positive (semi-)definite)?

 Granted the availability of a proper covariance matrix, this will ensure that both old and new latent variables satisfy the same linear relationships: (the population values of) their regression coefficients are the same, as are the coefficients of the simultaneous equation system, if any, and they have identical $R$-squares. In other words, consistent estimation methods like Lisrel and PLS, will remain consistent, and the estimators will not lose their asymptotic normality. The imposed non-normality will reveal itself, if it has an effect, only in changes in finite sample bias and instability.

Now consider the simplest possible non-linear factor model, where the one ‘inner equation’ reads:

$$\eta_3 = \gamma_1 \eta_1 + \gamma_2 \eta_2 + \gamma_{12} (\eta_1 \eta_2 - E(\eta_1 \eta_2)) + \zeta.$$  

(8)

Here $\zeta$ is independent of $\eta_1$ and $\eta_2$. All three latent variables have a zero

\(^4\)I am not sure this observation is worth anything, but the Hermite polynomials \(\{H_l(z_k)\}_{l=1}^{\infty}\) together with a constant, form an orthonormal basis of the Hilbert space of functions of $z_k$ with a finite variance. So roughly, by taking enough terms we can approximate in a least squares sense ‘any’ function of $z_k$ by Hermite polynomials. In particular, we can get any marginal distribution we want: if $\psi_k(\cdot)$ is the desired marginal distribution of $\tilde{\eta}_k$, then a least squares regression of $\Psi_k^{-1}(\Phi(z_k))$ on a sufficiently large number of Hermite polynomials comes arbitrarily close ($\Phi(\cdot)$ is the standard normal distribution).
mean and a unit variance. The regression coefficients $\gamma$ satisfy:

$$
\begin{bmatrix}
1 & E\eta_1 \eta_2 & E\eta_1^2 \eta_3 \\
E\eta_1 \eta_1 & 1 & E\eta_1^2 \eta_2 \\
E\eta_1 \eta_2 & E\eta_1 \eta_2 & E\eta_1^2 \eta_2 \cdot (E\eta_1 \eta_2)^2
\end{bmatrix}
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_{12}
\end{bmatrix}
= 
\begin{bmatrix}
E\eta_1 \eta_3 \\
E\eta_2 \eta_3 \\
E\eta_1 \eta_2 \eta_3
\end{bmatrix}.
$$

(9)

Under normality the covariance matrix of the regressors is ($\rho_{12} := E\eta_1 \eta_2$)

$$
\begin{bmatrix}
1 & \rho_{12} & 0 \\
\rho_{12} & 1 & 0 \\
0 & 0 & 1 + \rho_{12}^2
\end{bmatrix}
$$

(10)

and the righthand-side of (9) equals:

$$
\begin{bmatrix}
\gamma_1 + \gamma_2 \rho_{12} \\
\gamma_1 \rho_{12} + \gamma_2 \\
\gamma_{12} (1 + \rho_{12}^2)
\end{bmatrix}.
$$

(11)

In Klein & Muthén (2007) a similar model is analyzed. On page 660 and page 661 they briefly indicate how non-normality is introduced in a model with interactions, Study III. They write (slightly, immaterially adapted): For Study III, the data for the latent exogenous variables were generated non-normally using EQS with [skewness, kurtosis]=[−1.5, 4.0], [1.5, 5.0], [0.5, 5.0], respectively. The endogenous error variables were all simulated as normally distributed variables. (End of citation). This leads me to the third question:

3. Translated to our model (8), does this mean that the FVM approach is used on $\eta_1$ and $\eta_2$, so $\rho_{12}$ is kept fixed, but the other moments in the regression equation are ignored as restrictions?

If so, the underlying moment structure would be disrupted. I assume that $\tilde{\eta}_3$ is generated by $\gamma_1 \tilde{\eta}_1 + \gamma_2 \tilde{\eta}_2 + \gamma_{12}(\tilde{\eta}_1 \tilde{\eta}_2 - \rho_{12})$ plus a new independent (normal) residual, which generally will have a variance different from $E\zeta^2$. So the $R$-squared of the inner equation will be changed. Still assuming my reading is correct, the approach followed seems to deviate ‘strongly’ from the one followed in the linear case, where all moments relevant for the regression are maintained.

My fourth question is:

4. Would it make sense to try to honor the other moments as well?

If yes, we would need to add other terms to the ‘expansions’ of the $\tilde{\eta}_k$'s.
Since there are six additional equations to satisfy, one would add two additional terms to each $\tilde{\eta}_k$: $H_4(z_k)$ and $H_5(z_k)$, say. I do not kid myself that this will be easy, and I am not that much of a masochist, but still.

5. Perhaps it has been tried, or it imposes too many restrictions to yield a meaningful test of the importance of normality, or is simply known to fail?

The (tentatively) suggested approach here will be a real challenge for the situation where we have a full quadratic specification: we have 17 additional restrictions, requiring about 4 more terms per latent variable!

A final question:

6. What values for the skewness and kurtosis are deemed appropriate and on what grounds? (I do not suppose that researchers have made an effort to estimate the latent variables for (non-)linear models and study their distributional properties systematically?).


References


