**Polynomial SEM – uncomplete manuscript**

Reinhard Oldenburg

*Mathematics Department, Augsburg University, Augsburg, Germany*

March 2020

There are many approaches to nonlinear SEM but it seems that using Isserlis theorem to calculate the theoretical covariance matrix has not been used so far. This paper explores this approach.

Keywords: nonlinear polynomial equational modelling

**Introduction**

**Approach**

The approach in this paper is based on Isserlis’ theorem and straightforward.

**Theorem (Isserlis):** Assume $X_1, \ldots, X_n$ to be multivariate normally distributed and centred (e.g. expectation $E(X_i) = 0$) then the expectation of their product can be expressed in terms of covariances:

$$E(X_1 \cdot \ldots \cdot X_n) = \sum_{p \in \mathcal{P}_n^2} \prod_{(i,j) \in p} \text{cov}(X_i, X_j).$$

Where $\mathcal{P}_n^2$ is the set of all partition of $\{1, \ldots, n\}$ into disjoint subsets of size 2.

Now, the class of models that this paper deals with is specified. Similarly, like Bollen (1989, pp. 319) I assume that there are measurement models for $k$ exogenous ($\xi$) and $l$ endogenous ($\eta$) latent variables in terms of $m = m_1 + m_2$ observed variables $x, y$:

$$x = \Lambda_x \xi + \delta, y = \Lambda_y \eta + \epsilon.$$ 

And that $\xi, x, \delta$ are jointly normally distributed with zero expectation (this last assumption is not severe of course as the mean structure is rather trivial and can
approximately be fulfilled by subtracting the mean of observed data). No normality assumption is needed for \( y \). The structural model is given by a polynomial function and it involves another error term \( \zeta \):

\[
\eta = f(\xi) + \zeta .
\]

Note that \( \xi \) and \( \eta \) are vectors of random variable and thus \( f : \mathbb{R}^k \to \mathbb{R}^l \) consists of \( l \) real multivariate polynomial functions. Thus it can be written as

\[
f(\xi) = \sum_{(e_1, \ldots, e_k) \in \mathbb{N}_0^k} c(e_1, \ldots, e_k) \cdot \xi_1^{e_1} \cdot \ldots \cdot \xi_k^{e_k}
\]

where only finitely many of the \( c(e_1, \ldots, e_k) \in \mathbb{R}^l \) are nonzero.

All latent variables will be assumed to have zero expectation. If this is not the case, they can be replaced by the sum of a scalar parameter and a new, centered latent variable.

Furthermore, it is assumed that all components of error vectors are independent of each other and moreover

\[
cov(\delta, \epsilon) = cov(\xi, \delta) = cov(\xi, \epsilon) = cov(\zeta, \delta) = cov(\zeta, \epsilon) = 0.
\]

Under these assumptions the parameter implied covariance matrix can be calculated:

\[
\Sigma = \begin{pmatrix}
    cov(x, x') & cov(y, x') \\
    cov(y, x') & cov(y, y')
\end{pmatrix}
\]

The first entry is exactly calculated as in (Bollen 1989, pp. 323):

\[
cov(x, x') = E(xx') = E((\Lambda_x \xi + \delta)(\xi' \Lambda_x' + \delta'))
\]

\[
= \Lambda_x E(\xi \xi') \Lambda_x' + \Lambda_x E(\xi \delta') + E(\delta \xi') \Lambda_x' + E(\delta \delta')
\]

\[
= \Lambda_x E(\xi \xi') \Lambda_x' + E(\delta \delta')
\]

Here the last sum is diagonal because of the independency assumptions made above. Now, turn to the off-diagonal entry
\[
\text{cov}(y, x') = E \left( \left( (\Lambda_y f(\xi) + \zeta) + \epsilon \right) \cdot (\xi' \Lambda_x' + \delta') \right) = E \left( \left( (\Lambda_y f(\xi) + \zeta) \right) \cdot (\xi' \Lambda_x') \right) \\
= E(\Lambda_y f(\xi) \cdot \xi' \Lambda_x') = \Lambda_y E(f(\xi) \cdot \xi') \Lambda_x' \\
= \sum_{(e_1, \ldots, e_k) \in \mathbb{N}_6^k} \Lambda_y \cdot c_{e_1, \ldots, e_k} \cdot E(\xi_1^{e_1} \cdot \ldots \cdot \xi_k^{e_k} \cdot \xi') \Lambda_x'
\]

The last entry is not yet fully calculated, but it is clear that one needs only to evaluate the expectation on monomials of centred, normally distributed variables and therefore Isserlis’ theorem can be applied so that the result is a polynomial in covariances and parameters.

\[
\text{cov}(y, x') = E \left( \left( \Lambda_y f(\xi) + \zeta + \epsilon \right) \cdot \left( (f(\xi) + \zeta') \Lambda_y' + \epsilon' \right) \right) \\
= E \left( \left( \Lambda_y f(\xi) + \zeta \right) \cdot (f(\xi') + \zeta') \Lambda_y' \right) + E(\epsilon \epsilon') \\
= \Lambda_y E(f(\xi) f(\xi') + f(\xi) \zeta' + \zeta f(\xi')' + \zeta' \zeta') \Lambda_y' + E(\epsilon \epsilon') \\
= \Lambda_y E(f(\xi) f(\xi')' + f(\xi) \zeta' + \zeta f(\xi')') \Lambda_y' + \Lambda_y E(\zeta \zeta') \Lambda_y' + E(\epsilon \epsilon')
\]

Again, this is not yet fully calculated but it is obvious that linearity of the expectation and the polynomial structure of \( f \) allows this to be expanded so that Isserlis’ theorem can be applied.

Collecting the above results one arrives at:

**Theorem:** The model-based covariance matrix \( \Sigma \) of the polynomial SEM defined above consists of polynomials in the parameters of \( \Lambda_x, \Lambda_y \) as well as the variances of \( \epsilon, \zeta \) and variances and covariances of \( \xi \).

The process described above can easily be automated in computer algebra systems and this has been done in Mathematica.

The difference to linear SEM is that the entries in \( \Sigma \) are polynomials of higher degree while in the linear case the degree is linear in the variances and covariances.

Now, in principle any estimation method that minimizes some distance measure
between $\Sigma$ and $S$ can be applied, e.g. $F_{ULS} = \frac{1}{2} tr((S - \Sigma)^2)$ is a good choice as it does not depend on distributional assumptions. In contrast $F_{ML} = tr(S\Sigma^{-1}) + log|\Sigma| - log|S| - m$ will not lead to consistent estimations because nonlinearity of $f$ and normality of $x$ imply that $y$ will not be normal and hence $F_{ML}$ is not adequate. However, the $x$ part of the data is required to be multivariate normal and thus the following mixed strategy is obvious: The blocks of $\Sigma = \begin{pmatrix} \text{cov}(x, x') & \text{cov}(y, x') \\ \text{cov}(y, x') & \text{cov}(y, y') \end{pmatrix}$ are estimated with different methods: the whole objective function will be $F_{ML}(\text{cov}(x, x')) + 2F_{ULS}(\text{cov}(y, x')) + F_{ULS}(\text{cov}(y, y'))$. A sound theoretical basis has GLS estimation based on the theory developed by Browne (1984, eq. (3.4)), see also Bollen (1989, p. 426). Note, that what is mostly denoted by GLS is the special case of this theory for normal data but of course in the present case it is crucial to implement the general case.

As in the linear case the independency assumptions above can be relaxed somewhat by allowing some covariance to be non-zero but this is limited by the identification problem, of course.

**Case studies**

This paragraph reports on results of some practical applications. For some easy models (Oertzen 2020) the method brought perfect results. As a nontrivial test case I use Ganzach’s model as studied in Kelava & Brandt (2009), see Fig. 1.
Conclusion

The method presented is quite general as it can handle all polynomial SEM and yet the quality of estimations is quite good.

References


