Appendix A

In variance-based SEM, the path coefficients can be determined based on the construct score correlation matrix $R$, which for the model depicted in Figure 1 is of the following form ("cor" = correlation; "rel" = reliability):

$$
R = \begin{bmatrix}
1 & r_{12} & r_{13} \\
\rho_{12} & 1 & r_{23} \\
\rho_{13} & \rho_{23} & 1
\end{bmatrix}
$$

From $R$, we can extract the submatrix $R_x = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$ and the subvector $R_{xy} = \begin{bmatrix} r_{13} \\ r_{23} \end{bmatrix}$.

Under the assumption that the construct scores are standardized, the regression equation equals

$$
\hat{\beta} = \left[ \begin{array}{c}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{array} \right] = R_x^{-1} R_{xy} = \left[ \begin{array}{c}
1 \\
\rho_{12}
\end{array} \right] \left[ \begin{array}{c}
r_{13} \\
r_{23}
\end{array} \right] = \frac{1}{\det(R_x)} \left[ \begin{array}{c}
1 \\
\rho_{12}
\end{array} \right] \left[ \begin{array}{c}
1 \\
-\rho_{12}
\end{array} \right] = \frac{1}{1-\rho_{12}^2} \left[ \begin{array}{c}
r_{13} - \rho_{12} r_{23} \\
r_{23} - \rho_{12} r_{13}
\end{array} \right]
$$

Since the true $\beta_1$ equals zero, we can exploit that $\text{cor}(\xi_1, \eta) = \text{cor}(\xi_1, \xi_2) \cdot \text{cor}(\xi_2, \eta)$ and thus find the following result for the path $\hat{\beta}$ coefficient as estimated by variance-based SEM:
\[
\hat{\beta}_i = \frac{1}{1-r_{12}^2} \cdot (r_{13} - r_{12} r_{23}) \\
= \frac{\text{cor}(\xi_1, \eta) \cdot \text{rel}(\xi_1) \cdot \text{rel}(\eta) - \text{cor}(\xi_2, \xi_2) \cdot \sqrt{\text{rel}(\xi_1)} \cdot \sqrt{\text{rel}(\xi_2)} \cdot \text{cor}(\xi_2, \eta) \cdot \sqrt{\text{rel}(\xi_2)} \cdot \text{rel}(\eta)}{1 - \text{cor}^2(\xi_1, \xi_2)} \\
= \frac{\text{cor}(\xi_1, \xi_2) \cdot \text{cor}(\xi_2, \eta) \cdot \sqrt{\text{rel}(\xi_1)} \cdot \text{rel}(\eta) - \text{cor}(\xi_1, \xi_2) \cdot \sqrt{\text{rel}(\xi_1)} \cdot \text{rel}(\xi_2) \cdot \text{cor}(\xi_2, \eta) \cdot \sqrt{\text{rel}(\xi_2)} \cdot \text{rel}(\eta)}{1 - \text{cor}^2(\xi_1, \xi_2)} \\
= \frac{\text{cor}(\xi_1, \xi_2) \cdot \text{cor}(\xi_2, \eta) \cdot \sqrt{\text{rel}(\xi_1)} \cdot \text{rel}(\eta) \cdot (1 - \text{rel}(\xi_2))}{1 - \text{cor}^2(\xi_1, \xi_2)} \\
= \frac{- \text{cor}(\xi_1, \xi_2)}{1 - \text{cor}^2(\xi_1, \xi_2)} \cdot \text{cor}(\xi_2, \eta) \cdot \frac{1 - \text{rel}(\xi_2)}{\text{rel}(\xi_2)}
\]

Appendix B

A starting point for PLS-analyses is the so-called basic design, in essence a factor model. It is assumed that we have \( N \) i.i.d. column vectors of observed scores \( y_1, y_2, y_3, \ldots, y_N \) (i.e., we have \( N \) observations). We assume the usual standardization in PLS of each indicator, which means that each vector \( y_n \) for \( n = 1, 2, 3, \ldots, N \) has zero mean and its elements have unit variance. The vectors can be partitioned into \( M \) subvectors, \( M \geq 2 \), each with at least two components. For the \( i \)th subvector \( y_{in} \) of \( y_n \), we have

\[
y_{in} = \lambda_i \cdot \eta_{in} + e_{in}
\]

where the loading vector \( \lambda_i \) and the vector of measurement errors \( e_{in} \) have the same dimensions as \( y_{in} \), and the unobservable latent variable \( \eta_{in} \) is real-valued.\(^3\) For convenience, we make the sufficient but by no means necessary assumption that all components of all error vectors are mutually independent, and independent of all latent variables.

The latent variables have also zero mean and unit variance. The correlation between \( \eta_{in} \) and \( \eta_{jn} \) will be denoted by \( \rho_{ij} \). A particular set of easy implications is that the covariance matrix \( \Sigma_i \) of \( y_{in} \) can be written as

\[
\Sigma_i := E y_{in} y_{jn} = \rho_{ij} \lambda_i \lambda_j
\]

where the covariance matrix of the measurement errors \( \Theta_i \) is diagonal with non-negative diagonal elements, and we have for the covariance between \( y_{in} \) and \( y_{jn} \)

\(^1\)To keep the paper self-contained, we collect here the main algebraic results underlying PLSc (for a more elaborate explication, see Dijkstra 2010). Note that there are many ways to correct for inconsistency and the one we use here is quite possibly the simplest one. See Dijkstra (2013) for alternatives and extensions. Any combination of PLS with one of the alternatives could legitimately be called PLSc.

\(^2\)The use of i.i.d. (independent and identically distributed) signifies that the data are seen as a random sample from a population. This is for conceptual convenience only. Together with the classical laws of large numbers and the central limit theorem (Cramér 1946), this entails the consistency and asymptotic normality properties we need. We can make much more general assumptions, but the statistical subtleties would be a distraction and would not lead to different results (CAN-estimators).

\(^3\)We follow exactly the specification of Joreskög (1969). Please note that variables are arranged per rows, and observations per columns. Moreover, note that, as opposed to our custom in the main body of the paper, it pays here to use a subscript for the blocks.
\[
\sum_{j} \equiv E_{j}^{\prime} y_{j} = \rho_{y}^{\lambda}, \lambda_{j}^{T}
\] (3)

The covariance matrix of each \( y_{j} \) will be denoted by \( \Sigma \), and the sample covariance matrix by \( S \). The sample counterparts of \( \Sigma \) and \( \Sigma_{i} \) are denoted by \( S_{i} \) and \( S_{i} \), respectively. Since we assume that the sample data are standardized before being analyzed, we have \( S_{i} = \frac{1}{N} \sum_{j=1}^{N} y_{j} y_{j}^{T} \).

Note that the assumptions made so far entail that the sample counterparts are consistent and jointly asymptotically normal estimators of the theoretical variance and covariance matrices (Cramér 1946, Chapter 28).

PLS features a number of iterative fixed-point algorithms, of which we select one, the so-called Mode A algorithm. This is, in general, numerically the most stable algorithm.\(^4\) As a rule, it converges from arbitrary starting vectors, and it is usually very fast. The outcome is an estimated weight vector \( \hat{w}_{i} \), with typical subvector \( \hat{w}_{j} \) of the same dimensions as \( y_{i} \). With these weights, sample proxies are defined for the latent variables: \( \hat{\eta}_{j} \equiv \hat{w}_{j}^{T} y_{i} \) with the customary normalization of a unit sampling variance, so \( \frac{1}{N} \sum_{i=1}^{N} (\hat{w}_{j}^{T} y_{i})^{2} = \hat{w}_{j}^{T} S_{i} \hat{w}_{j} = 1 \). In Wold’s PLS approach the \( \hat{\eta}_{j} \)’s replace the unobserved latent variables, and loadings and structural parameters are estimated using ordinary least squares.\(^5\) For Mode A, we have for each \( i \) \( \approx \) stands for “proportional to”

\[
\hat{w}_{j} \approx \sum_{j \in C(i)} \text{sign}_{ij} \cdot S_{j} \hat{w}_{j}
\] (4)

Here, \( \text{sign}_{ij} \) is the sign of the sample correlation between \( \hat{\eta}_{j} \) and \( \hat{\eta}_{j} \), and \( C(i) \) is a set of indices of latent variables. Traditionally, \( C(i) \) contains the indices of latent variables adjacent to \( \hat{\eta}_{j} \) (i.e., the indices of latent variables that appear on the other side of the structural or path equations in which \( \hat{\eta}_{j} \) appears). Clearly, \( \hat{w}_{j} \) is obtained by a regression of the indicators \( y_{i} \), on the sign-weighted sum: \( \sum_{j \in C(i)} \text{sign}_{ij} \cdot \hat{\eta}_{j} \). There are other versions (with correlation weights, for example); this is one of the very simplest, and it is the original one (see Wold, 1982). There is little motivation in the PLS literature for the coefficients of \( S \hat{w} \),\(^6\) but the particular choice can be shown to be irrelevant for the probability limits of the estimators. The algorithm takes an arbitrary starting vector, and then basically follows the sequence of regressions for each \( i \), each time inserting updates when available (or after each full round; the precise implementation is not important).

Dijkstra (1981, 2010) has shown that the PLS modes converge with a probability tending to one when the sample size tends to infinity, for essentially arbitrary starting vectors. Moreover, the weight vectors that satisfy the fixed-point equations, are locally continuously differentiable functions of the sample covariance matrix of \( y \). They, as well as other estimators that depend smoothly on the weight vectors and \( S \), are therefore jointly asymptotically normal.

Let us denote the probability limit of \( \hat{w}_{i} \), \( \text{plim} \hat{w}_{i} \), by \( \bar{w}_{i} \). We can get it from the equation for \( \hat{w}_{i} \) by substitution of \( \Sigma \) for \( S \).

\[
\bar{w}_{j} \approx \sum_{j \in C(i)} \text{sign}_{ij} \cdot \sum_{j \in C(i)} \bar{w}_{j} = \sum_{j \in C(i)} \text{sign}_{ij} \cdot \rho_{y} \lambda_{j} \bar{w}_{j} = \lambda_{j} \sum_{j \in C(i)} \text{sign}_{ij} \cdot \rho_{y} \lambda_{j} \bar{w}_{j}
\] (5)

Since \( \lambda_{j} \bar{w}_{j} \) is a scalar, and all terms in the sum have the vector \( \lambda_{j} \) in common, it is clear that \( \bar{w}_{j} \approx \lambda_{j} \). We must have \( \bar{w}_{i}^{T} \sum \bar{w}_{i} = 1 \), so

\[
\bar{w}_{j} = \frac{\lambda_{j}}{\sqrt{\sum_{i} \lambda_{i}}}
\] (6)

---

\(^4\)This is because Mode A ignores collinearity between the observed variable predictors of the proxy, while Mode B takes account of that collinearity and thus must find the inverse of the covariance matrix of the predictions, which itself is sometimes unstable.

\(^5\)Other estimators like 2SLS are also possible.

\(^6\)In favor of the sign-weights, one could argue that in sufficiently large samples \( \text{sign}_{ij} \cdot S_{j} \hat{w}_{i} \) is approximately equal to \( \lambda_{j} \cdot \rho_{y} \left( \lambda_{j}^{T} \bar{w}_{j} \right) \), where the term in brackets measures the (positive) correlation between \( \hat{\eta}_{j} \) and its proxy; see below for results that help justify this claim. So the tighter the connection between \( \eta_{j} \) and the better \( \bar{\eta}_{j} \) can be measured, the more important \( \bar{\eta}_{j} \) is in determining \( \bar{w}_{j} \).
We conclude that PLS, Mode A, produces estimated weight vectors that tend to vectors proportional to the true loadings. One would like to have a simple estimate for the proportionality factor. We propose here as in Dijkstra (1981, 2010, 2011) to define

\[
\hat{\lambda}_{ii} = \frac{\hat{w}_i^T \left( S_{ii} - \text{diag}(S_{ii}) \right) \hat{w}_i}{\hat{w}_i^T \left( \hat{w}_i \hat{w}_i^T - \text{diag}(\hat{w}_i \hat{w}_i^T) \right) \hat{w}_i}
\]

as a function of \( c \) and obtain

\[
\hat{c}_i = \sqrt{\frac{\lambda_{ii} \Sigma_{ii}^2}{\lambda_{ii}^2 + \Sigma_{ii}^2}} = \text{plim} \hat{\lambda}_{ii} = \text{plim} \hat{c}_i
\]

Now, in particular

\[
\text{plim} \hat{\lambda}_{ii} = \text{plim} \hat{c}_i = \text{plim} \hat{\lambda}_{ii} = \text{plim} \hat{c}_i = \frac{\lambda_{ii} \Sigma_{ii}^2}{\lambda_{ii}^2 + \Sigma_{ii}^2}
\]

It will be useful to define a population proxy \( \eta_{in} \) by \( \eta_{in} = \hat{w}_i \hat{y}_{in} \). Clearly, the squared correlation between a population proxy and its corresponding latent variable is

\[
R^2(\eta_{in}, \eta_i) = \left( \hat{w}_i^T \hat{\lambda}_{ii} \right)^2
\]

which equals

\[
\left( \hat{\lambda}_{ii}^2 \hat{\lambda}_{ii} \right)^2 + \lambda_{ii} \Sigma_{ii} \hat{\lambda}_{ii} = \frac{\lambda_{ii} \Sigma_{ii}^2}{\lambda_{ii}^2 + \Sigma_{ii}^2}
\]

With a large number of high quality indicators, this correlation can be close to one ("consistency at large" in PLS parlance). A trivially deduced but important algebraic relationship is

\[
R^2(\eta_{in}, \eta_i) = \left( \hat{w}_i^T \Sigma_{ii} \hat{w}_i \right)^2 = \rho_{\hat{w}_i}^2 \cdot R^2(\eta_{in}, \eta_i) \cdot R^2(\eta_{in}, \eta_i)
\]

indicating that the PLS proxies will tend to underestimate the squared correlations between the latent variables. In fact, one can show that this is true for multiple correlations as well (see Dijkstra 2010). Also note that

\[
R^2(\eta_{in}, \eta_i) = \left( \hat{w}_i^T \hat{\lambda}_{ii} \right)^2 = \left( \hat{w}_i^T \left( \hat{w}_i \cdot \hat{c}_i \right) \right)^2 = \left( \hat{w}_i^T \hat{w}_i \right)^2 \cdot \hat{c}_i^2
\]

so that we can estimate the (squared) quality of the proxies consistently by

\[
R^2(\eta_{in}, \eta_i) = \left( \hat{w}_i^T \hat{\lambda}_{ii} \right)^2 = \left( \hat{w}_i^T \left( \hat{w}_i \cdot \hat{c}_i \right) \right)^2 = \left( \hat{w}_i^T \hat{w}_i \right)^2 \cdot \hat{c}_i^2
\]

This assumes that the measurement errors within a block are uncorrelated, the basic design. If we know that some errors are correlated or we have doubts about them, we can delete the items from the difference to be minimized.
\[ \hat{R}^2(\eta, \eta) := \left( \hat{w}_i^T \hat{w}_i \right)^2 \cdot \hat{e}_i^2 \]  

(15)

Moreover, with

\[ \hat{R}^2(\eta_i, \eta_j) := \left( \hat{w}_i^T S_{ij} \hat{w}_j \right)^2 \cdot \hat{e}_i^2 \]  

(16)

we can estimate the correlations between the latent variables consistently (see Equation 13)

\[ \hat{\rho}_i := \sqrt{\frac{\hat{R}^2(\eta, \eta)}{\hat{R}^2(\eta_i, \eta_i) \cdot \hat{R}^2(\eta_i, \eta_i)}} \]  

(17)

We close this appendix with four observations:

1. Standard PLS software for Mode A will produce all the necessary, simple ingredients for consistent estimation.

2. The approach can be and has been extended to Mode B (Dijkstra 1981, 2010, 2011), but since Mode A is numerically more stable and faster than Mode B, it has first priority.

3. In the main body of this paper, we used \( \rho(\eta_i) \) for \( R(\eta_i, \eta_i) \).

4. In the main body of this paper, we used (without subscript \( i \))

\[ \rho_{\eta,i} := \left( \hat{w}_i^T \hat{w}_i \right)^2 \cdot \frac{\hat{w}_i^T \left( S_{ji} - \text{diag}(S_{ji}) \right) \hat{w}_j}{\hat{w}_i^T \hat{w}_i \cdot \hat{w}_j^T \left( \hat{w}_i^T \hat{w}_i - \text{diag}(\hat{w}_i^T \hat{w}_i) \right) \hat{w}_j} \]  

(18)

This is just \( \hat{R}^2(\eta_i, \eta_i) \).

References


