In H. Latan, J.F. Hair, & R. Noonan (Eds.), Partial least squares path modeling: Basic concepts, methodological issues, and applications (2nd ed.). Cham, Switzerland: Springer

Statistical and Psychometric Properties of Three Weighting Schemes of the PLS-SEM Methodology

Ke-Hai Yuan and Zhiyong Zhang
University of Notre Dame

June 5, 2022
Revised, September 21, 2022
Second revision, November 24, 2022
Third revision, December 22, 2022
Fourth revision, January 5, 2022

This work was supported by a grant from the Department of Education (R305D210023) and in part by Grant 31971029 from the Natural Science Foundation of China. However, the contents of the study do not necessarily represent the policy of the funding agencies, and endorsement by the Federal Government should not be assumed.

Abstract

Structural equation modeling (SEM) is a widely used technique for studies involving latent constructs. While covariance-based SEM (CB-SEM) permits estimating the regression relationship among latent constructs, the parameters governing this relationship do not apply to that among the scored values of the constructs, which are needed for prediction, classification and/or diagnosis of individuals/participants. In contrast, the partial-least-squares approach to SEM (PLS-SEM) first obtains weighted composites for each case and then estimates the structural relationship among the composites. Consequently, PLS-SEM is a preferred method in predicting and/or classifying individuals. Nevertheless, properties of PLS-SEM still depend on how the composites are formulated. Herman Wold proposed to use mode A to compute the scores for constructs with reflective indicators. However, Yuan and Deng recently showed that composites under mode B enjoy better psychometric properties. The authors thus proposed a structured transformation from mode A to mode B, denoted as mode B_A. This chapter further studies properties of the three modes of PLS-SEM. Analytical and numerical results show that 1) Mode A does not possess any solid statistical or psychometric properties, 2) Mode B possesses good theoretical properties but is over sensitive to sampling errors, and 3) Mode B_A possesses good theoretical properties as well as numerical stability. The performances of the three modes are also illustrated with two real data examples.

1. Introduction

Structural equation modeling (SEM) and path analysis with weighted composites are among the most widely used methods in social and behavioral sciences. The two distinguished classes of methods are integrated in the so-called partial-least-squares approach to structural equation modeling (PLS-SEM). Based on the measurement and structural model, PLS-SEM first obtains weighted composites[1] to act as proxies to the latent variables, and then to estimate the structural model by regression analysis with the weighted composites. To differentiate the conventional SEM methodology from PLS-SEM, the former is often called covariance-based SEM (CB-SEM). The most well-known feature of CB-SEM is its capability of separating measurement errors from latent constructs. This feature facilitates consistent parameter estimates as well as test statistics and fit indexes for evaluating the goodness-of-fit of the overall model structure. In contrast, PLS-SEM or regression analysis with weighted composites directly estimates the relationship among the scored-values of the composites and has the strength of maximizing the predictive roles of the exogenous variables on the endogenous variables according to the principle of LS regression (Boardman, Hui & Wold, 1981; Cho et al., 2022a; Hair et al., 2017; Wold, 1980). Still, the properties of PLS-SEM closely depend on how the weights of the composites are computed. Wold (1980, 1982) proposed two algorithms to compute the weights, termed as modes A and B, respectively. Yuan and Deng (2021) introduced a new mode, termed as mode $\mathrm{B_A}$, and section 4 of this chapter provides a detailed description of this mode. The purpose of this chapter is to systematically study the properties of the three weighting schemes, including statistical properties of scale invariance and scale-inverse equivariance when the scales of the observed variables change, as well as the psychometric properties of measurement reliability of the resulting composites. These properties will be obtained analytically and illustrated via numerical and real data examples. To fully understand the performances of the three modes in operation, we will also discuss the sensitivity of the weights with imperfect data.

It is well-known that the scales of latent variables have to be fixed in order for the SEM model to be identified (see e.g., Loehlin & Beaujean, 2017). This is typically done by fixing either the variance of the latent variable at 1.0 or one of the loadings of its indicators at 1.0. The two choices are arbitrary and so is the value of 1.0. Although the overall model structure of the observed variables remains the same regardless of how the scale of each latent variable is fixed, the values of the parameters in the measurement and structural models depend on

[1] Throughout the chapter, a weighted composite or composite-score is a weighted sum of the observed values of items that are designed to measure a latent construct.

these choices. This implies that particular population values of the model parameters under CB-SEM are artificial. In parallel, the scales of the composites under PLS-SEM also need to be determined. This is typically done by fixing the variance of each composite at 1.0. While such a choice has become the norm in the field, one can also choose a different set of values for the scales of the composites without affecting any substantive aspect of the model. In particular, we can always choose the scales of the latent variables or those of the composites so that the two methods have identical values of path coefficients (see e.g., Devlieger, Mayer & Rosseel, 2016; Skrondal & Laake, 2001; Yuan & Deng, 2021). Consequently, we will not compare parameter estimates of PLS-SEM against those of CB-SEM in this chapter. Instead, we will focus on the properties of the weights and the resulting composites under the three modes of the PLS-SEM methodology. The properties of the resulting parameter estimates under each mode will also be examined when the scales of the observed variables change. Because the formulations of composites and the efficiency of parameter estimates under PLS-SEM are totally determined by the weights of the items, our study of weights not only clarifies the pros and cons of the different modes of the methodology but also facilitates better understanding of other approaches of path analysis with weighted composites.

Although CB-SEM has the advantage in yielding consistent estimates of path coefficients, the parameters are for characterizing the relationship among latent variables that represent the population distribution, and all individuals/participants are equivalent under such a relationship. In practice when scored values of composites are used for prediction or diagnosis, individuals are no longer equivalent. An individual with greater scores is expected to perform better on the criterion variable, and such a relationship is directly characterized by the regression model with the composite-scores. In particular, the regression model with LS estimates still yields the best (i.e., smallest mean-squared error) linear unbiased predictor for a future value even when predictors contain measurement errors (see Fuller, 1987, p. 75). However, not all weighted composites are equivalent in prediction. The values of R^2 as well as the relative errors of the estimated regression coefficients depend on the measurement reliabilities of the composites (Yuan & Fang, 2022), which further depend on the formulation of the weights.

By focusing on models with reflective indicators, we will discuss the following aspects of the weights of composites in this chapter: The measurement reliability of the resulting composites; the reactions of the weights, the composites and the resulting regression coefficients to scale change of the observed variables; sensitivity of the weights to model misspecification, negative weights and negative estimates of error variances (Heywood cases). Numerical and real-data examples will be used in the analysis. Because PLS-SEM is relatively new to researchers in social and behavioral sciences, we will give a brief introduction to the methodology by pointing out some of its distinctive features.

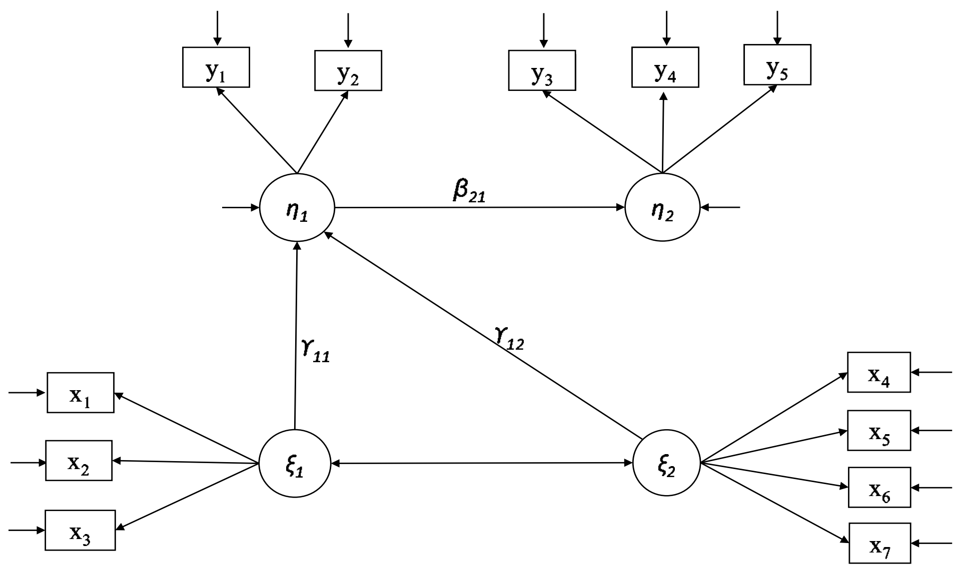
The reason for us to mainly consider models with reflective indicators is because weights of composites for formative indicators are not determined by the indicators themselves but by their relationships with indicators of other latent variables or composites (Treiblmaier, Bentler & Mair 2011). Also, with formative indicators, the concept of measurement reliability may not apply since the indicators may not contain measurement errors nor do they need to share anything in common. In addition, our discussion of model misspecification is via the traditional factor model, not the composite models as described in Dijkstra (2017), Cho and Choi (2020), and Hwang et al. (2020). Furthermore, the analytical results on weights obtained by Dijkstra (1983) only hold for models with reflective indicators, and they will be further studied in this chapter. But our results on scale invariance and scale-inverse equivariance also apply to formative indicators and composite models, as will be further discussed in the concluding section.

A useful preface for following the development of this chapter is that PLS-SEM mode B does not have to stick to models with formative indicators, although such a match was recommended by Wold (1980, 1982). Actually, the analytical results by both Dijkstra (1983) and Schneeweiss (1993) include applying mode B to models with reflective indicators. This chapter presents additional theoretical advantages of mode B for models with reflective indicators, and they are shared by mode $\mathrm{B_A}$ at the level of population.

2. Two distinctive features of PLS-SEM and the environmental variable

While PLS-SEM is essentially path analysis with weighted composites (Hair et al., 2017; Henseler, 2021), the method is self-contained with its own algorithms for computing the composites and conducting parameter estimation of the structural model. Consider the path diagram in Figure 1, where there are four latent variables ξ_1, ξ_2, η_1, η_2 and twelve reflective indicators, and the indicators of each latent variable are referred to as a block. There are no correlated errors nor cross loadings in Figure 1. That is, each indicator only loads on the single latent variable of its block. Such a property is commonly termed as *unidimensionality* in measurement, which is a distinctive feature of PLS-SEM. Although unidimensionality is not necessary under CB-SEM, the property is very desirable because it facilitates theory testing and development as well as assessment and evaluation (e.g., reliability, validity, interpretability) (see e.g., Anderson & Gerbing, 1984). However, the model will be misspecified when either cross loadings or error covariances exist in the population. Even if CB-SEM permits the inclusion of cross loadings and/or correlated errors, model misspecification[2] cannot be avoided in practice (MacCallum, 2003), which will cause biased parameter estimates under both CB-SEM and PLS-SEM. We will discuss the effects of misspecified models on the different weighting schemes of the PLS-SEM methodology in a later section, based on recent results by Yuan, Wen and Tang (2022). Effects of misspecified models on parameter estimates under CB-SEM can be found in Yuan, Marshall and Bentler (2003).

Figure 1. A model with four latent variables and twelve indicators.



Another distinctive feature of PLS-SEM is the way in *counting direct connections*, which is needed in the algorithms for computing the weights. In Figure 1, ξ_1 is directly connected with η_1; ξ_2 is directly connected with η_1; η_1 is directly connected with ξ_1, ξ_2, and η_2; and η_2 is directly connected with η_1. However, the two-way arrow between ξ_1 and ξ_2 is not considered as a direct connection under PLS-SEM. Such a way of counting connections generates weighted composites by which the corresponding indicators partially maximize their predictive relationships (Boardman et al., 1981). The maximum relationship is operated by LS regression via the so-called *environmental variable*, which plays the role of being the representative of the directly connected constructs (Schneeweiss, 1993). In Figure 1, the

[2]Regardless of reflective or formative indicators, a model is misspecified whenever the model-implied covariance matrix does not equal the population covariance matrix of all the involved indicators. Modeling truly formative indicators as reflective or vice versa is expected to cause a discrepancy between the model-implied covariance matrix and its population counterpart.

environmental variable of ξ_1 is $\bar{\xi}_1 = c_{\xi_1\eta_1}\eta_1$, where $c_{\xi_1\eta_1}$ can be the sign of the correlation between ξ_1 and η_1 or the correlation itself. Similarly, the environmental variable of η_1 is $\bar{\eta}_1 = c_{\eta_1\xi_1}\xi_1 + c_{\eta_1\xi_2}\xi_2 + c_{\eta_1\eta_2}\eta_2$, where each c can be either the sign of the correlation (termed as the centroid scheme) or the value of the correlation itself (termed as the factorial scheme). Thus, the environmental variable of a focal latent variable is a linear combination of the latent variables that are directly connected to the focal latent variable. In operation, when the latent variables are approximated by scored values of composites, environment variables will become the corresponding linear combinations of the composites. As to be described in the following section, weights of indicators are obtained by LS regression, which maximizes the linear relationship of each indicator with the corresponding environmental variable.

As noted earlier, composites do not have natural scales. They need to be assigned and are of an arbitrary nature. In PLS-SEM, this is done by scaling the weights for each block of indicators so that the resulting composite has a variance of 1.0. Also, for variables with a single connection, the value of the coefficient (c) in the environmental variable is cancelled due to standardization.

3. PLS-SEM modes A and B

PLS-SEM methodology consists of two stages. Weights of composites are computed in the first stage via an iterative process, and regression analysis with the weighted composites is conducted in the second stage. Consider the model in Figure 1, let the initial composite for each latent variable be the simple average of its block of indicators, and followed by a standardization so that the sample variance of each composite is 1.0. The corresponding environmental variables are obtained when the constructs are replaced by their composites, where the coefficients c are also computed according to the correlations of the corresponding composites. Under mode A, weights of x_1 to x_3 are updated by the LS regression coefficient of each of the indicators on the environmental variable $\bar{\xi}_1$ (simple regression), and weights of x_1 to x_3 under mode B are updated by the LS regression coefficients of the environmental variable $\bar{\xi}_1$ on x_1, x_2 and x_3 (multiple regression). Weights of indicators in the other blocks are updated in parallel by the LS regression coefficients via the corresponding environmental variables. The updated weights of each block are proportionally rescaled so that the corresponding composite has a sample variance of 1.0. The environmental variables are then updated via the updated composites, which completes a cycle of iteration in estimating the weights. The iteration process continues with the updated environmental variables until the weights for all the blocks of indicators are stabilized, and the corresponding composites are consequently obtained. These weighted composites represent the constructs in conducting path analysis at stage 2. Although stage 2 can be done via fitting the sample covariance matrix of the composites by the structural model using different methods for covariance structure analysis (e.g., ML, LS, GLS), PLS-SEM uses separate LS regression to estimate the path coefficients for each endogenous construct (Wold, 1980, 1982), which is easy to carry out.

Conventionally, mode A has been recommended for models with reflective indicators and mode B for models with formative indicators (Wold, 1980, 1982). While such recommendations are followed in software[3] and textbooks (Hair et al., 2017), they are based on intuition rather than justified by statistical or psychometric theory. In particular, Yuan and Deng (2021) showed that, when applying mode B to reflective indicators, the method is asymptotically equivalent to regression analysis using Bartlett-factor scores (BFS). They also showed that regression-factor scores[4] are equivalent to Bartlett-factor scores in conducting regression analysis. Note that BFSs attain the maximum reliability among all weighted composites (see e.g., Bentler, 1968; Yuan & Bentler, 2002). The results in Yuan and Deng (2021) imply that composites under PLS-SEM mode B attain the maximum reliability asymptotically. Because more reliable composites correspond to more efficient estimates of regression coefficients and greater R^2 values (Cochran, 1970; Yuan & Fang, 2022), mode B is theoretically more preferred than mode A for estimating models with reflective indicators.

We use numerical examples to show the theoretical advantage of mode B. For a block of indicators, let the variance of the latent factor be 1.0, $\boldsymbol{\lambda}$ be the vector of factor loadings, and $\boldsymbol{\Psi}$ be the diagonal matrix of error variances. Then the covariance matrix of the block of indicators is given by $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Psi}$. The equivalence between BFS regression and PLS-SEM model B was based on the results in Dijkstra (1983) and Schneeweiss (1993), who showed that the weight vector under mode A is proportional to $\boldsymbol{\lambda}$ and that under mode B is proportional to $\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$. Let's use $\mathbf{w}_a$ and $\mathbf{w}_b$ to denote the weights under the two modes. The results in Dijkstra (1983), Schneeweiss (1993) and Yuan and Deng (2021) imply that $\mathbf{w}_a = c_a\boldsymbol{\lambda}$ and $\mathbf{w}_b = c_b\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}$, where c_a and c_b are scalars so that the corresponding composites have a variance of 1.0. Table 1 contains four examples and each has three items. The population factor loadings λ_j and error variances ψ_{jj} for the first three examples are exact,

[3]A reviewer noted that software SmartPLS automatically assumes that indicators are reflective when mode A is chosen, and are formative when mode B is chosen.

[4]Note that regression-factor scores can be separately computed for each block of indicators or collectively computed for all the blocks of indicators. It is the separately computed regression-factor scores that are equivalent/proportional to the BFSs, which remain the same whether collectively or separately computed.

Table 1. Reliabilities of three composites with examples: ω is the reliability of equally weighted composite, ρ_A and ρ_B are the reliabilities of composites under PLS-SEM modes A and B, respectively.

		population θ		reliability			
	variable	λ_j	ψ_{jj}	ρ_j	ω	ρ_A	ρ_B
Example 1	x_1	0.400	0.840	0.160	0.597	0.666	0.697
	x_2	0.500	0.750	0.250			
	x_3	0.800	0.360	0.640			
Example 2	x_1	0.350	0.8775	0.123	0.562	0.695	0.746
	x_2	0.400	0.8400	0.160			
	x_3	0.850	0.2775	0.723			
Example 3	x_1	0.400	0.500	0.242	0.579	0.570	0.581
	x_2	0.500	0.700	0.263			
	x_3	0.800	0.900	0.416			
Example 4	x_1	0.492	0.758	0.242	0.567	0.578	0.581
	x_2	0.513	0.737	0.263			
	x_3	0.645	0.584	0.416			

Note. Items in Example 4 are obtained by the standardization of those in Example 3. The population values of the parameters for Examples 1 to 3 are exact while those for Example 4 are rounded.

while those for the fourth example are rounded. The reliabilities are computed according to the population factor loadings and error variances, where ρ_j is the reliability of the jth item, ω is the reliability of the equally-weighted composite (EWC), ρ_A and ρ_B are respectively the reliabilities of composites under PLS-SEM modes A and B. With the variance of the latent variable in each example being 1.0, we have $\rho_j = \lambda_j^2/(\lambda_j^2 + \psi_{jj})$. The formulas for computing ρ_A and ρ_B were given by Yuan and Deng (2021), and that for computing ω was given by McDonald (1999), and was often called Dillon-Goldstein's ρ in the PLS-SEM literature (see Vinzi et al., 2010).

In Example 1, the variance of each item is 1.0. The EWC has a reliability of .597 but the reliability of x_3 is .640. All the items in Example 2 also have variances at 1.0, and the reliability of x_3 is .723. Both the EWC and the weighted composite under PLS-SEM mode A are less reliable than the single item x_3. In Example 3, the items are not standardized. The weighted composite under mode A is less reliable than the EWC. The items in Example 4 are obtained by standardizing those in Example 3. The results of Examples 3 and 4 show that ω and ρ_A are scale dependent, while ρ_B is scale invariant. We will present analytical results regarding the properties of different modes under scale-transformation in a later section.

Based on Monte Carlo results Henseler et al. (2014, p. 190) stated "PLS Mode A outperforms the best indicator across all model constellations, providing support for the capability of PLS to reduce measurement error." Results in Table 1 suggest that the comparison of reliabilities among composites under mode A, EWCs, and individual indicators depends on conditions. When a block of indicators contains an item with rather larger reliability than the rest of the items, then it is hard for the composite under mode A or the EWC to be more reliable, as is the case in Example 2. When factor loadings and error variances are close to proportional, the reliability of the EWC is close to maximum while composites under PLS-SEM mode A can be less reliable. This is the case in Example 3. In general, if items with larger loadings have even larger error variances, then composites under mode A will be less reliable than equally weighted composites (see Yuan, Wen & Tang, 2020). But the composite under mode B reaches the maximum reliability regardless of the reliabilities of the individual items, although sampling errors may have a negative effect on the estimated weights of the mode. We will further discuss the sensitivity of mode B to sampling and model specification errors in the following sections.

Note that for simplicity we only considered models with a single construct in this section. The conclusions with the four examples also hold for models with more latent constructs.

4. PLS-SEM mode $\mathrm{B_A}$

While PLS-SEM mode B yields composites with maximum reliability at the level of population, Dijkstra and Henseler (2015a) noted that mode A is numerically more stable. An example with a real dataset in Yuan and Deng (2021) showed that weights of some individual items under PLS-SEM mode B are negative. For the same dataset, all the individual weights under PLS-SEM mode A and all the factor loadings under CB-SEM are positive. Because negative weights are not logically acceptable for positively worded items, PLS-SEM mode B has problems in operation. In order to have weighted composites that enjoy the statistical/psychometric properties of mode B while the method also performs as stable as mode A numerically, Yuan and Deng (2021) proposed a procedure to transform the weights under mode A to weights that are asymptotically equivalent to those under mode B, or a transformed mode $\mathrm{B_A}$.

Let $\hat{\mathbf{w}}_a = (\hat{w}_{a1}, \hat{w}_{a2}, \ldots, \hat{w}_{ap})'$ be the estimated weights under mode A and $\mathbf{S}$ be the sample covariance matrix of a given block that has p indicators. The transformed mode is obtained via fitting the sample covariance matrix $\mathbf{S}$ by the one factor model

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \phi \hat{\mathbf{w}}_a \hat{\mathbf{w}}_a' + \mathbf{\Psi}, \tag{1}$$

where ϕ plays the role of factor variance, $\hat{\mathbf{w}}_a$ plays the role of factor loadings, $\mathbf{\Psi} = \mathrm{diag}(\psi_{11}, \psi_{22},$

$\ldots, \psi_{pp})$ is a diagonal matrix of the unexplained variances, and $\boldsymbol{\theta} = (\phi, \psi_{11}, \psi_{22}, \ldots, \psi_{pp})'$ contains $(p+1)$ free parameters. The model in equation (1) can be estimated by normal-distribution-based maximum likelihood (NML) or least-squares (LS). Appendix E of Yuan and Deng (2021) contains the development of the LS solutions, which are given by

$$\hat{\phi} = (\hat{\mathbf{w}}_a'\mathbf{S}\hat{\mathbf{w}}_a - \sum_{j=1}^{p} \hat{w}_{aj}s_{jj})/[(\sum_{j=1}^{p} \hat{w}_{aj})^2 - \sum_{j=1}^{p} \hat{w}_{aj}^4] \text{ and } \hat{\psi}_{jj} = s_{jj} - \hat{\phi}\hat{w}_{aj}^2. \tag{2}$$

With the estimates in (2), the estimated weights $\hat{\mathbf{w}}_{b_a} = (\hat{w}_{b_a1}, \hat{w}_{b_a2}, \ldots, \hat{w}_{b_ap})'$ under PLS-SEM mode $\mathrm{B_A}$ are give by

$$\hat{w}_{b_aj} = c_{b_a}\hat{\psi}_{jj}^{-1}\hat{w}_{aj}, \ j = 1, 2, \ldots, p,$$

where c_{b_a} is a scalar so that the weighted composite under model $\mathrm{B_A}$ has a sample variance of 1.0.

Because $\hat{\mathbf{w}}_a$ converges in probability to $\mathbf{w}_a = c_a\boldsymbol{\lambda}$ and $\mathbf{S}$ converges to $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Psi}$, the LS estimates $\hat{\psi}_{jj}$ in equation (2) converge to ψ_{jj}. Consequently, $\hat{\mathbf{w}}_{b_a}$ converges to $\mathbf{w}_b = c_b\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}$. Thus, the weights under PLS-SEM mode $\mathrm{B_A}$ are asymptotically equivalent to those under PLS-SEM mode B and those of the corresponding Bartlett-factor score. The resulting composites under mode $\mathrm{B_A}$ also enjoy the same theoretical properties as composites under mode B or the Bartlett-factor scores.

As for estimating any other factor models in factor analysis, the estimates $\hat{\psi}_{jj}$ in equation (2) can be negative (Heywood case). In such a case, we can replace the negative estimate by $\tilde{\psi}_{jj} = .05$ (or another small number) and adjust the value of $\hat{w}_{aj}$ via

$$\tilde{w}_{aj}^2 = (s_{jj} - \tilde{\psi}_{jj})/\hat{\phi},$$

yielding

$$\tilde{w}_{b_aj} = c_{b_a}\tilde{\psi}_{jj}^{-1}\tilde{w}_{aj}.$$

The adjusted $\tilde{w}_{aj}$ is to keep $\hat{\phi}\tilde{w}_{aj}^2 + \tilde{\psi}_{jj} = s_{jj}$. One can also only adjust the value of $\hat{\psi}_{jj} < 0$ to a small positive number without adjusting the value of $\hat{\mathbf{w}}_a$.

With correctly specified models, Heywood cases are mostly due to a small sample size together with small population values of ψ_{jj}. Model misspecification and/or data contamination are also responsible for Heywood cases in practice. Thus, negative estimates of error variances can offer additional information about the model, the data, and/or the population, and they should be regarded as an opportunity rather than a bad luck.

The literature of PLS-SEM repeatedly claims that the methodology has solved the issue of negative estimates of error variances (e.g., Chin, 1998; Henseler, 2021, p. 162). This is because the estimand of error variance under PLS-SEM is different from that under CB-SEM, and the former includes both measurement and prediction errors. In contrast, the ψ_{jj}s in equation (1) only represent the variances of measurement errors, assuming no unique factors nor systematic errors in the model. Regardless of whether there exist measurement errors, LS regression never yields a negative estimate of prediction-error variance.

5. Scale invariance and scale-inverse equivariance

We have shown in section 3 that composites under mode A may not be as reliable as a single indicator. In this section we will examine two additional statistical properties of the three modes, *scale invariance* and *scale-inverse equivariance*. These are fundamental because they describe how parameters react when the scales of the observed variables change. In particular, we will show that PLS-SEM mode A is scale dependent, and the use of standardized variables is to hide the issue of scale dependency of the method rather than having the issue solved. In contrast, weights under PLS-SEM mode B and mode $\mathrm{B_A}$ are scale-inverse equivariant and the resulting composites and regression coefficients are scale invariant. For simplicity, we will present the analytical results by a one-factor model and numerical illustration by a two-factor model. The results also hold for more complex models, as will be shown in section 7 via real data examples.

5.1 Analytical results

Let $\mathbf{x}$ be a vector of mean-centered random variables representing a block of indicators. Suppose $\mathbf{x}$ follows a one-factor model with $\mathbf{x} = \boldsymbol{\lambda}\xi + \boldsymbol{\varepsilon}$ and

$$\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Psi},$$

where $\mathrm{Var}(\xi) = 1$ for model identification and $\boldsymbol{\Psi} = \mathrm{Cov}(\boldsymbol{\varepsilon})$ is a diagonal matrix. Dijkstra (1983) and Schneeweiss (1993) showed that weights under PLS-SEM mode A are proportional to $\boldsymbol{\lambda}$. That is,

$$\mathbf{w}_{ax} = c_{ax}\boldsymbol{\lambda}, \ \text{ with } \ c_{ax} = (\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda})^{-1/2}. \tag{3}$$

The corresponding composite is given by

$$\hat{\xi}_{ax} = \mathbf{w}_{ax}'\mathbf{x} = c_{ax}\boldsymbol{\lambda}'\mathbf{x}. \tag{4}$$

Let $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_p)$ be a diagonal matrix with $d_j > 0$. When the variables in $\mathbf{x}$ are scaled according to $\mathbf{y} = \mathbf{D}\mathbf{x}$, then the covariance matrix of $\mathbf{y}$ and weights under

PLS-SEM mode A respectively become

$$\mathrm{Cov}(\mathbf{y}) = \mathbf{D\Sigma D} = \mathbf{D}\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathbf{D} + \mathbf{D\Psi D} \ \text{ and } \ \mathbf{w}_{ay} = c_{ay}\mathbf{D}\boldsymbol{\lambda}, \tag{5}$$

where $c_{ay} = (\boldsymbol{\lambda}'\mathbf{D}^2\boldsymbol{\Sigma}\mathbf{D}^2\boldsymbol{\lambda})^{-1/2}$. The composite corresponding to $\mathbf{y}$ is given by

$$\hat{\xi}_{ay} = \mathbf{w}'_{ay}\mathbf{y} = c_{ay}\boldsymbol{\lambda}'\mathbf{D}\mathbf{y} = c_{ay}\boldsymbol{\lambda}'\mathbf{D}^2\mathbf{x}. \tag{6}$$

Thus, both the weights $\mathbf{w}_a$ and the composite $\hat{\xi}_a$ depend on the scales of the indicators. Equation (5) indicates that $\mathbf{w}_a$ is *scale equivariant*. That is, each weight is transformed the same way as the corresponding indicator. However, equation (6) indicates that the values of the d_js are squared in the resulting $\hat{\xi}_{ay}$, which is very undesirable.

The procedure of standardization corresponds to $\mathbf{D} = \mathrm{diag}(1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_p)$, where σ_j is the standard deviation of the jth indicator of the block. The results in equations (3) to (6) show that working with standardized variables does not make mode A a scale-free method.

Dijkstra (1983) and Schneeweiss (1993) also showed that, when applying PLS-SEM mode B to a correctly specified model with reflective indicators, the resulting $\mathbf{w}_b$ is proportional to $\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$. Via an analytical expression for $\boldsymbol{\Sigma}^{-1}$, Yuan and Deng (2021) showed that $\mathbf{w}_b$ can be equivalently expressed as

$$\mathbf{w}_{bx} = c_{bx}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}, \ \text{ with } \ c_{bx} = (\boldsymbol{\lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda})^{-1/2}.$$

The corresponding composite is given by $\hat{\xi}_{bx} = \mathbf{w}'_{bx}\mathbf{x} = c_{bx}\boldsymbol{\lambda}'\boldsymbol{\Psi}^{-1}\mathbf{x}$. Since the error variances of $\mathbf{y}$ are given by the diagonal of $\mathbf{D\Psi D}$, applying the above formula to the scale-transformed variables we have

$$\mathbf{w}_{by} = c_{by}(\mathbf{D\Psi D})^{-1}\mathbf{D}\boldsymbol{\lambda} = c_{by}\mathbf{D}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda} \ \text{ and } \ \hat{\xi}_{by} = \mathbf{w}'_{by}\mathbf{y} = c_{by}(\mathbf{D}^{-1}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda})'\mathbf{y} = c_{by}\boldsymbol{\lambda}'\boldsymbol{\Psi}^{-1}\mathbf{x},$$

where c_{by} is a scalar such that $\mathrm{Var}(\hat{\xi}_{by}) = 1$. Clearly, $c_{by} = c_{bx}$ and $\hat{\xi}_{by} = \hat{\xi}_{bx}$. Thus, the weights in $\mathbf{w}_b$ are *scale-inverse equivariant* and the composite $\hat{\xi}_b$ is *scale invariant*. These properties imply that PLS-SEM mode B will yield the same regression coefficients whether standardized variables or raw measurements are used in the analysis.

At the population level, the weight vector corresponding to the mode $\mathrm{B_A}$ is given by

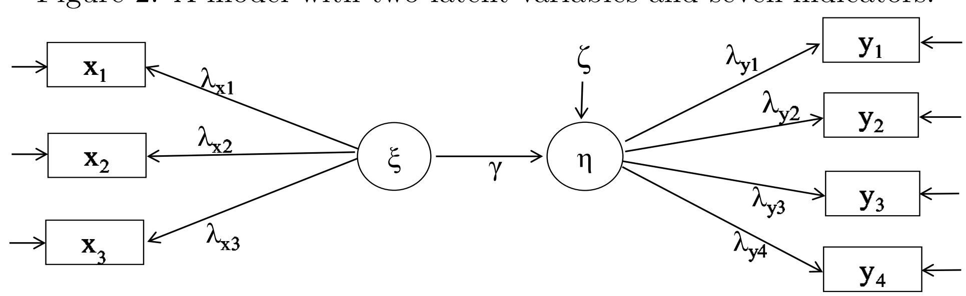
$$\mathbf{w}_{b_a} = c_{b_a}\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda},$$

where c_{b_a} is a constant such that $\mathbf{w}'_{b_a}\boldsymbol{\Sigma}\mathbf{w}_{b_a} = 1$, which implies $c_{b_a} = c_b$. When the variables in $\mathbf{x}$ are rescaled according to $\mathbf{y} = \mathbf{D}\mathbf{x}$, the resulting $\boldsymbol{\Sigma}_y = \mathrm{Cov}(\mathbf{y})$ and the weight vector for $\mathbf{y}$ under mode A are given by equation (5). Then the population counterpart of the $\mathbf{\Psi}$ in equation (1) becomes $\mathbf{\Psi}_y = \mathbf{D\Psi D}$. Thus, the weight vector under the mode $\mathrm{B_A}$ for the transformed variables is given by $\mathbf{w}_{b_a} = c_{b_a}\mathbf{\Psi}_y^{-1}(\mathbf{D}\boldsymbol{\lambda}) = c_{b_a}\mathbf{D}^{-1}\mathbf{\Psi}^{-1}\boldsymbol{\lambda}$, which is the same as $\mathbf{w}_b$. Therefore, weights under the mode $\mathrm{B_A}$ are scale-inverse equivariant, and the corresponding composite $\hat{\xi}_{b_a}$ as well as the regression coefficients are scale invariant.

5.2 Numerical results

We use an example to illustrate the different properties of the three modes numerically. Figure 2 represents a model with 2 latent constructs and 7 indicators. Let the population values of the parameters be given by $\boldsymbol{\lambda}_x = (0.80, 1.00, 1.20)'$, $\boldsymbol{\lambda}_y = (1.00, 1.20, 0.80, 1.50)'$, $\gamma = .70$, $\phi = \mathrm{Var}(\xi) = 1.00$, and $\sigma_\zeta^2 = \mathrm{Var}(\zeta) = .40$. The matrices of the error variances are $\mathbf{\Psi}_x = \mathrm{diag}(.55, .60, .40)$ and $\mathbf{\Psi}_y = \mathrm{diag}(.58, .65, .60, .40)$. The resulting population covariance matrix $\mathbf{\Sigma}$ of the 7 indicators is given in Table A1 of appendix A.

Figure 2. A model with two latent variables and seven indicators.



Applying the PLS-SEM modes A, $\mathrm{B_A}$, and B to this population covariance matrix yields three population weight vectors, respectively. They are reported in the left panel of Table 2(a) and are denoted as $\mathbf{w}_\sigma$. When each of the 7 variables is standardized, we have a corresponding correlation matrix $\mathbf{P}$. Applying the three modes of PLS-SEM to this correlation matrix yields three different weight vectors, and they are in the middle panel of Table 2(a). The three vectors in the right panel are obtained by applying $\mathbf{D}_\sigma\mathbf{w}_\sigma$ to each of the vector on the left panel, where $\mathbf{D}_\sigma$ is the diagonal matrix consisting of the population standard deviations of the 7 indicators. For both modes B and $\mathrm{B_A}$, there exists $\mathbf{D}_\sigma\mathbf{w}_\sigma = \mathbf{w}_\rho$ or $\mathbf{w}_\sigma = \mathbf{D}_\sigma^{-1}\mathbf{w}_\rho$, verifying the scale-inverse equivariance properties of the two modes. However, the weight vectors under mode A clearly do not enjoy such a property. In addition, modes B and $\mathrm{B_A}$ yield identical weight vectors in each case, which verifies that mode $\mathrm{B_A}$ is asymptotically equivalent to mode B, and both yield composites with maximal reliability.

Table 2(b) contains the parameters of the regression models $\eta_w = \gamma_w\xi_w + e_w$ corresponding

Table 2. (a) Population values of the weights of PLS-SEM modes A, B_A and B under the analyses of the covariances σ_{ij} ($\mathbf{w}_\sigma$), the correlations ρ_{ij} ($\mathbf{w}_\rho$), and by transformation ($\mathbf{D}_\sigma \mathbf{w}_\sigma$).

	Modeling $\mathbf{\Sigma}$ ($\mathbf{w}_\sigma$)			Modeling $\mathbf{P}$ ($\mathbf{w}_\rho$)			$\mathbf{D}_\sigma \mathbf{w}_\sigma$		
weight	A	B_A	B	A	B_A	B	A	B_A	B
w_{x1}	0.241	0.210	0.210	0.348	0.230	0.230	0.263	0.230	0.230
w_{x2}	0.301	0.241	0.241	0.375	0.305	0.305	0.381	0.305	0.305
w_{x3}	0.362	0.434	0.434	0.420	0.589	0.589	0.490	0.589	0.589
w_{y1}	0.189	0.163	0.163	0.283	0.198	0.198	0.229	0.198	0.198
w_{y2}	0.226	0.175	0.175	0.296	0.243	0.243	0.315	0.243	0.243
w_{y3}	0.151	0.126	0.126	0.254	0.137	0.137	0.163	0.137	0.137
w_{y4}	0.283	0.356	0.356	0.332	0.551	0.551	0.439	0.551	0.551

(b) Population values for the regression models corresponding to the three modes of PLS-SEM under the analyses of the covariances σ_{ij} ($\boldsymbol{\theta}_\sigma$), and the correlations ρ_{ij} ($\boldsymbol{\theta}_\rho$).

	$\boldsymbol{\theta}_\sigma$			$\boldsymbol{\theta}_\rho$		
parameter	A	B_A	B	A	B_A	B
γ_w	0.653	0.656	0.656	0.646	0.656	0.656
σ_w^2	0.573	0.569	0.569	0.583	0.569	0.569
$\mathcal{R}_w^2$	0.427	0.431	0.431	0.417	0.431	0.431

Note: $\mathbf{P}$ is the population correlation matrix; $\mathbf{D}_\sigma$ is a diagonal matrix whose diagonal elements are the square root of the diagonal elements of $\mathbf{\Sigma}$; and $\mathcal{R}_w^2$ is the population counterpart of R^2.

to the three modes under the population covariances and correlations, where η_w and ξ_w are the weighted composites and γ_w is the corresponding regression coefficient. The results in Table 2(b) show that the regression coefficient γ_w, the error variance $\sigma_w^2 = \text{Var}(e_w)$, and the coefficient of determination $\mathcal{R}_w^2$ (population R-square) are scale invariant under both modes B and B_A. But the parameters under mode A do not possess such a property.

5.3 Sample results

For the population covariance matrix $\mathbf{\Sigma}$ (Table A1) that generated the results in Table 2, a sample of size N=100 is drawn from the normal population $N(\mathbf{0}, \mathbf{\Sigma})$. The SAS IML program that generated the sample as well as the 100 × 15 data matrix are available at `www3.nd.edu\~kyuan\PLS-SEM_property`. The code of the SAS program is also provided in appendix B for reference, and the sample covariance matrix of this sample is given in Table A2 of appendix A. Table 3 contains the results of applying modes A, B_A and B to this sample. Parallel to Table 2, the estimated weights corresponding to modeling the sample covariances ($\mathbf{S}$) and sample correlations ($\mathbf{R}$) are denoted by $\hat{\mathbf{w}}_s$ and $\hat{\mathbf{w}}_r$, respectively; and the transformed weights are obtained by $\mathbf{D}_s \hat{\mathbf{w}}_s$, where $\mathbf{D}_s$ is a diagonal matrix whose diagonal elements are given by the square roots of those of $\mathbf{S}$. The estimates of the model parameters for $\hat{\eta}_w = \gamma_w \hat{\xi}_w + e_w$ are also reported in Table 3.

Table 3. (a) Estimated weights of PLS-SEM modes A, $\mathrm{B_A}$ and B under the analyses of the sample covariances s_{ij} ($\hat{\mathbf{w}}_s$), the sample correlations r_{ij} ($\hat{\mathbf{w}}_r$), and by transformation ($\mathbf{D}_s\hat{\mathbf{w}}_s$).

	Modeling $\mathbf{S}$ ($\hat{\mathbf{w}}_s$)			Modeling $\mathbf{R}$ ($\hat{\mathbf{w}}_r$)			$\mathbf{D}_s\hat{\mathbf{w}}_s$		
weight	A	$\mathrm{B_A}$	B	A	$\mathrm{B_A}$	B	A	$\mathrm{B_A}$	B
$\hat{w}_{x1}$	0.245	0.192	0.211	0.345	0.226	0.262	0.304	0.239	0.262
$\hat{w}_{x2}$	0.256	0.217	0.194	0.357	0.280	0.246	0.325	0.275	0.246
$\hat{w}_{x3}$	0.315	0.390	0.393	0.385	0.570	0.568	0.456	0.563	0.568
$\hat{w}_{y1}$	0.126	0.034	0.200	0.199	0.030	0.239	0.151	0.040	0.239
$\hat{w}_{y2}$	0.239	0.123	-0.223	0.336	0.122	-0.297	0.318	0.163	-0.297
$\hat{w}_{y3}$	0.180	0.059	-0.126	0.271	0.056	-0.158	0.224	0.074	-0.158
$\hat{w}_{y4}$	0.320	0.528	-0.526	0.395	0.856	-0.801	0.487	0.805	-0.801

(b) Estimates for the regression models corresponding to the three modes of PLS-SEM under the analyses of the sample covariances s_{ij} (θ_s) and the sample correlations r_{ij} (θ_r).

	$\hat{\boldsymbol{\theta}}_s$			$\hat{\boldsymbol{\theta}}_r$		
parameter est	A	$\mathrm{B_A}$	B	A	$\mathrm{B_A}$	B
$\hat{\gamma}_w$	0.649	0.667	-0.686	0.634	0.666	-0.686
$\hat{\sigma}_w^2$	0.584	0.560	0.534	0.604	0.562	0.534
R_w^2	0.422	0.446	0.471	0.402	0.443	0.471

Note: $\mathbf{D}_s$ is a diagonal matrix whose diagonal elements are the square root of the diagonal elements of $\mathbf{S}$.

As expected, estimates under PLS-SEM mode A do not possess the property of scale-inverse invariance nor scale invariance when the scales of the indicators change. The estimated weights under mode B are scale-inverse equivariant and the estimated regression parameters are scale invariant. But three of the four weights for the y-indicators by mode B are negative. Note that we let the first element of each weight vector be positive in iteratively computing the weights. If we had let $\hat{w}_{y1}$ under mode B be negative, then $\hat{w}_{y2}$, $\hat{w}_{y3}$ and $\hat{w}_{y4}$ would be positive as would $\hat{\gamma}_w$.

The results in Table 3 imply that weights under mode $\mathrm{B_A}$ do not possess the property of scale-inverse equivariance nor the parameter estimates possess that of scale invariance with finite samples. This is because the sampling errors in the estimated weight $\hat{\mathbf{w}}_a$ and those in $\mathbf{S}$ affect the estimates of the error variances in equation (2). In particular, weights under mode $\mathrm{B_A}$ are obtained by transforming those under mode A. When the model in equation (1) does not fit the sample covariance matrix $\mathbf{S}$ perfectly, the effect of sampling errors is carried to weights under mode $\mathrm{B_A}$. However, compared to the different weights under mode A, the elements of $\hat{\mathbf{w}}_r$ and $\mathbf{D}_s\hat{\mathbf{w}}_s$ under mode $\mathrm{B_A}$ are much closer to each other. This is because weights under mode B_A are asymptotically scale-inverse equivariant and estimates of regression parameters are asymptotically scale invariant.

The values of the R^2 in Table 2 and Table 3 show that, while modes B and B_A are asymptotically equivalent, the former yields a greater R^2 value at the sample level. In contrast, mode A has the smallest R^2 in both the population and the sample.

Because our focus is PLS-SEM, we did not formally discuss the invariance or equivariance properties of regression analysis with equally-weighted composites (EWC) in this section. Although EWC regression is widely used in practice and was recommended against PLS-SEM by some authors (e.g., Rönkkö et al., 2022), the results of EWC regression are neither scale invariant nor scale equivariant nor scale-inverse equivariant. The reliabilities of EWCs are also scale dependent. Parallel to PLS-SEM mode A, the use of standardized items or standardized composites under EWC regression is to avoid addressing the issues of scale dependency of the method. We will use an example to illustrate such facts in a later section.

6. Sensitivity of weights to misspecified models

Results in the previous section and those in Yuan and Deng (2021) indicate that mode B can yield negative weights although the mode has theoretical advantages. This section discusses the sensitivity of different modes to model misspecification for the purpose of better understanding the empirical behaviors of the different modes. Detailed analysis on the sensitivity of weights to model misspecification was given in Yuan, Wen and Tang (2022), and we only briefly summarize the results here.

A latent-variable model typically includes measurement model and structural model. Yuan, Wen and Tang (2022) only considered misspecification in the measurement model, mostly because the measurement model under PLS-SEM is rather restricted. In contrast, the structural model under PLS-SEM can be specified as saturated. Thus, we also only consider misspecified measurement models, which may mistakenly exclude three types of parameters: (1) within-block error covariances, (2) between-block error covariances, and (3) cross loadings. Our discussion will be for the three types of misspecification. For simplicity, we will only discuss the case with two latent variables. But the conclusions equally hold for models with more latent constructs, as to be illustrated via an empirical example in the next section. Interested readers are referred to Yuan, Wen and Tang (2022) for a comprehensive study.

Consider the model in Figure 2, which has seven measurement errors and seven factor loadings. Each error might be correlated with the errors of its own block or of the other block. Each indicator might also load on the latent variable of the other block. When the PLS-SEM mode is clear from the context, we will use $\mathbf{w}_x$ and $\mathbf{w}_y$ to represent the vectors of weights corresponding to the blocks $\mathbf{x}$ and $\mathbf{y}$, respectively. Since the scales of the weighted composites are arbitrary, we will focus on the relative change of the weights using their counterparts under the correctly specified model as the reference. In particular, we will regard the weights as not-affected if they are proportional to their counterparts under the correctly specified model (i.e., no cross loadings nor correlated errors in the population). Because the order of the indicators within each block is arbitrary, the conclusions obtained for a particular indicator also apply to other indicators of the same nature within the same block.

6.1 PLS-SEM Mode A

Within-block error covariance

When x_1 and x_2 in Figure 2 have correlated errors, neither the weight vector $\mathbf{w}_x$ nor the weight vector $\mathbf{w}_y$ under the mode A of PLS-SEM is affected. Each vector is still proportional to that of the correctly specified model, i.e., the vector of factor loadings of the respective block. Parallel results hold when y_2 and y_3 in Figure 2 have correlated errors. That is, within-block correlated errors do not affect the weights under PLS-SEM mode A.

Between-block error covariance

When x_3 and y_4 in Figure 2 have correlated errors, only the individual weights of x_3 and y_4 are affected. The other elements of $\mathbf{w}_x$ and $\mathbf{w}_y$ under mode A are still proportional to their respective factor loadings. If there is a 3rd latent variable with a block of reflective indicators, the weight vector for this block will not be affected by the error correlations between the other two blocks.

Cross loading

In Figure 2 when x_1 has a nonzero loading on η, only the weight for x_1 in $\mathbf{w}_x$ is affected. The weights for the other indicators in the block $\mathbf{x}$ are still proportional to their respective factor loadings. The weight vector $\mathbf{w}_y$ is not affected. In parallel, the existence of a cross loading of y_4 on ξ in Figure 2 only affects the weight of y_4 in $\mathbf{w}_y$. The other elements of $\mathbf{w}_y$ as well as the whole vector $\mathbf{w}_x$ are still proportional to their respective factor loadings.

6.2 PLS-SEM Mode B

Within-block error covariance

When y_3 and y_4 in Figure 2 have correlated errors, the weights for y_3 and y_4 in $\mathbf{w}_y$ are affected. The other weights within the block are still proportional to their respective factor loadings. The weight vector $\mathbf{w}_x$ is not affected. Parallel results hold when errors in the block $\mathbf{x}$ are correlated. That is, within-block error covariances only affect the individual weights of the involved items. They do not affect the weights for the other items within the same block nor any of the weights of a different block.

Between-block error covariance

When x_3 and y_4 in Figure 2 have correlated errors, only the individual weights of x_3 and y_4 are affected. The other elements of $\mathbf{w}_x$ and $\mathbf{w}_y$ under mode B are still proportional to their counterparts under the correctly specified model (i.e., the vector of factor loading multiplied by the precision matrix of the block). If there is a 3rd latent variable with a block of reflective indicators, the weight vector for this block will not be affected by the error covariances between the other two blocks.

Cross loading

When x_1 has a nonzero loading on η in Figure 2, only the weight for x_1 in $\mathbf{w}_x$ is affected. The weights for the other indicators in the block $\mathbf{x}$ are still proportional to their counterparts under the correctly specified model. The weight vector $\mathbf{w}_y$ is still proportional to the vector of factor loadings multiplied by the precision matrix of the block $\mathbf{y}$. In parallel, the existence of a cross loading of y_4 on ξ in Figure 2 only affects the weight of y_4 in $\mathbf{w}_y$, and the other elements of $\mathbf{w}_y$ are still proportional to their counterparts under a correctly specified model, and so is the whole vector $\mathbf{w}_x$.

6.3 PLS-SEM Mode $\mathrm{B_A}$

Within-block error covariance

When x_1 and x_2 in Figure 2 have correlated errors, the one-factor model in equation (1) is misspecified. To account for the need of fitting the covariances of the indicators, the factor variance (the parameter ϕ) in equation (1) has to take a different value from that of the correctly specified model, and so do the error variances. Consequently, any error covariances within the block $\mathbf{x}$ will affect all the elements of $\mathbf{w}_x$ under the mode $\mathrm{B_A}$. However, error covariances within the block $\mathbf{x}$ do not affect the weight vector $\mathbf{w}_y$. Parallel results hold when $\mathbf{y}$ has within-block error covariances. That is, within-block error covariances only affect the weight vector of the corresponding block. They do not affect the weights of a different block.

Between-block error covariance

When x_3 and y_4 in Figure 2 have correlated errors, all the elements of $\mathbf{w}_x$ and $\mathbf{w}_y$ are affected under the mode $\mathrm{B_A}$. This is because a change in a single element of the weight vector under mode A will affect the communalities of all the indicators in $\mathbf{x}$ via the change of the factor variance (the parameter ϕ) when equation (1) is estimated. They together cause all the elements of $\mathbf{w}_x$ and $\mathbf{w}_y$ under mode $\mathrm{B_A}$ to change. However, if there is a 3rd latent variable with a block of reflective indicators, the weight vector for this block will not be affected by the error covariances between the other two blocks.

Cross loading

When x_1 has a nonzero loading on η in Figure 2, the change in weight of x_1 under mode A will affect the values of the communalities of all the indicators in $\mathbf{x}$ via the change of the factor variance (the parameter ϕ in equation 1). They further cause changes in all the elements of $\mathbf{w}_x$ under the mode $\mathrm{B_A}$. However, the vector $\mathbf{w}_y$ under mode $\mathrm{B_A}$ is not affected by the existence of a cross loading of x_1 on η.

We only discussed whether the weights will change or remain intact when an extra association in the population exists. The size of the change also varies between the different modes. In particular, mode B tends to have the largest change and is very sensitive to the existence of cross loadings, although more elements of the weights under mode $\mathrm{B_A}$ are affected (Yuan, Wen & Tang, 2022). Because model misspecification and sampling errors are empirically confounded, the results in this section explain why some weights under mode B in Table 3 are negative although the two-factor model is correct.

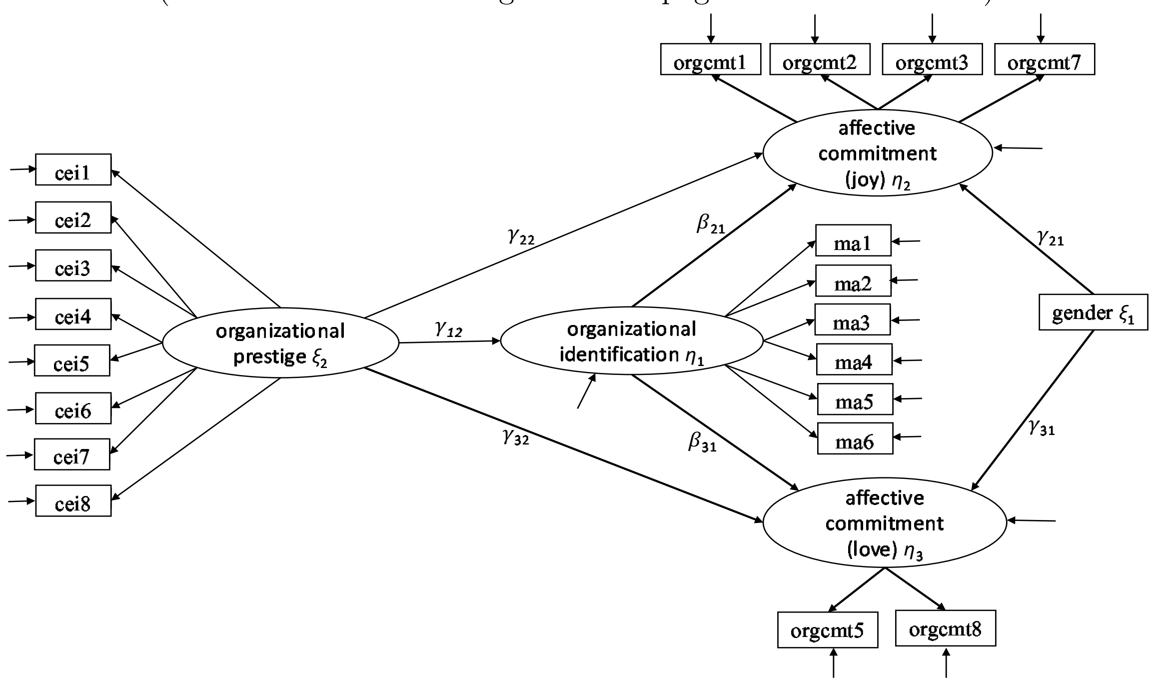
7. Two real data examples

This section contains two examples, and the purpose is to illustrate the properties of the PLS-SEM methodology presented in the previous sections, including the size of measurement reliability, scale invariance and equivariance of weights and reliability coefficients, and the sensitivity of weights. The data and model for the first example are from a PLS-SEM textbook; and the dataset for the second example is from a classical textbook of multivariate statistics that has been used to illustrate various developments in SEM.

Example 5.

The path diagram in Figure 1.3 of Henseler (2021, p. 6) presented a model with 21 reflective indicators and 4 latent constructs. One of the 21 indicators is gender, playing the role of a covariate (ξ_1). The other 20 indicators include eight organizational prestige indicators for a latent construct ξ_2 (organizational prestige), six organizational identification indicators for a latent construct η_1 (organizational identification), four affective commitment indicators for a latent construct η_2 (joy), and two affective commitment indicators for a latent construct η_3 (love). The wording of the indicators can be found in Table 6.1 of Henseler (2021, p. 135). According to the table, this organization-culture dataset has 22 variables from 305 participants, and they were part of a larger survey among South Korean employees conducted and reported by Bergami and Bagozzi (2000). One of the indicators (orgcmt6) in Table 6.1 of Henseler (2021) is not used in his Figure 1.3. Among the 305 participants, there are 157 male employees and 148 female employees. For easy reference, a path diagram for the model is given in Figure 3 of this chapter, which has 7 path coefficients: γ_{21}, γ_{31}, γ_{12}, γ_{22}, γ_{32}, β_{21}, and β_{31}. Note that the labels for joy and love in Figure 1.3 of Henseler (2021) are switched from his Table 6.1, and the labels in our Figure 3 adopted those in his Table 6.1. Our purpose with this real-data example is to use the textbook model to illustrate the discussed properties of the three modes of the PLS-SEM methodology.

Figure 3. A model with a covariate (gender), four latent variables and twenty indicators (the same model as in Figure 1.3 on page 6 of Henseler 2021).



Because both LS regression and the NML method for CB-SEM are strongly influenced by data contamination and/or outlying observations, we first check the distribution properties of this organization-culture dataset. Since gender is unlikely contaminated, it is excluded from the examination. The standardized Mardia's (1970) multivariate kurtosis of the remaining 20 variables is $M_s = 40.481$, which is highly significant when compared against the standard normal distribution. We thus use a robust method to control the heavy tails of the data, and choose a Huber-type M-estimator for the purpose (Huber, 1981). This can be carried out via a robust transformation procedure (Yuan, Chan & Bentler, 2000). Note that the mechanism of robustness with Huber-type M-estimator is to give smaller weights to outlying cases in estimating the means and covariances. The purpose is to get more efficient parameter estimates or numerically more stable results in the analysis (Yuan & Gomer, 2021). Let's denote the robust method via Huber-type weights as H(φ) when cases whose Mahalanobis distances (d_M^2) are greater than the $1 - \varphi$ quantile of χ^2_{20} are downweighted. We started with H(.05), however, the standardized multivariate kurtosis of the transformed sample is still highly significant ($M_s = 9.112$). Following the recommendation in Yuan and Gomer (2021), we continued to increase the value of φ by checking the corresponding standardized multivariate kurtosis until $\varphi = .25$, and the corresponding value of the M_s is -0.099. Our analysis and comparison below are to fit the model in Figure 3 to the transformed sample by H(.25), where gender is not subject to the transformation.

Clearly, the model in Figure 3 can be estimated under both PLS-SEM and CB-SEM. Although our main interest in this chapter is to study psychometric and statistical properties of PLS-SEM, these properties are characterized via the model implied covariance matrix. Thus, we will first estimate the model under CB-SEM via the NML method to obtain this covariance matrix. Considering that PLS-SEM researchers may not have much interest in results under CB-SEM, these are put in Appendix C of this chapter. Instead, we will include the results of path analysis with EWCs, since the method lacks the fundamental properties of a principled method but was favored against PLS-SEM.

For each of the four latent variables (ξ_2, η_1, η_2, η_3), fixing their first factor loading at 1.0 for model identification, the CB-SEM model has 48 free parameters. Fitting the model in Figure 3 to the robustly transformed sample by NML results in $T_{ml} = 613.756$, corresponding to a p-value that is essentially 0 when referred to χ^2_{183}. Fit indices (RMSEA=.088, CFI=.871) also indicate that the model in Figure 3 fits the data marginally[5] according to the established norms (Hu & Bentler, 1999). This is not unusual when working with real data. The goodness-of-fit can be improved by modifying the model in Figure 3 (e.g., correlated errors and/or cross loadings). But our purpose is to use the textbook example with the original model to illustrate the fundamental properties of the PLS methodology, as stated earlier.

The estimates of the factor loadings (λ), factor variances (ϕ), path coefficients (γ, β), measurement error variances (ψ) and prediction error variances (σ^2_ζ) of the CB-SEM model are reported in Table C1, where the standard errors (SE) and z-statistics are computed according to the ML (i.e., NML) method in EQS (Bentler, 2006). All the parameter estimates

[5] When the CB-SEM model is fitted to the original (not robustly transformed) data by NML, the results are $T_{ml} = 669.183$, RMSEA=.093, and CFI=.848. The estimate of γ_{32} corresponds to a z-statistic at -1.840.

are statistically significant at the level of .05.

Insert Table 4 about here

We next apply PLS-SEM modes A, $\mathrm{B_A}$ and B to estimate the model in Figure 3. When working with the unstandardized variables (modeling the sample covariance matrix $\mathbf{S}$) of the robustly transformed sample the 21 estimated weights by each method are given in the left panel of Table 4, and those with the standardized variables (modeling the sample correlation matrix $\mathbf{R}$) are in the middle panel. The right panel of Table 4 contains the results of $\mathbf{D}_s\hat{\mathbf{w}}_s$, where $\mathbf{D}_s$ is a diagonal matrix whose diagonal elements are the square root of the diagonal elements of $\mathbf{S}$, and $\hat{\mathbf{w}}_s$ is the vector of weights on the left panel of the table. Clearly, the estimated weights by mode B under modeling $\mathbf{R}$ are identical to those under $\mathbf{D}_s\hat{\mathbf{w}}_s$, verifying the property of scale-inverse equivariance of the mode. In contrast, the estimated weights by mode A do not enjoy such a property. The weights under mode $\mathrm{B_A}$ in Table 4 are not strictly scale-inverse equivariant either, due to sampling errors and model misspecification. But the values of $\hat{w}_{b_a}$ under $\mathbf{D}_s\hat{\mathbf{w}}_s$ are rather close to those under modeling $\mathbf{R}$.

Note that the first value of the estimated weights for each block of indicators in Table 4 is set as positive. However, five of the eight values of $\hat{w}_b$ for the indicators of ξ_2 are negative. This shows the sensitivity of mode B to model misspecification/sampling errors, although the weights are scale-inverse equivariant. In contrast, weights under modes A and $\mathrm{B_A}$ are all positive, conforming with the expectation in formulating composites.

Insert Table 5 about here

Table 5 contains the estimated path coefficients of the structural model in Figure 3, where EWC regression is also included for comparison purpose. Clearly, the results by mode B remains the same whether working with the unstandardized variables (modeling $\mathbf{S}$) or the standardized variables (modeling $\mathbf{R}$). The results by mode $\mathrm{B_A}$ under modeling $\mathbf{R}$ are almost identical to those under modeling $\mathbf{S}$, whereas those by PLS-SEM mode A clearly depend on the scales of the observed indicators. The results of EWC regression in Table 5 are even more affected by the scales of the observed indicators than those of PLS-SEM mode A, indicating that the former is not a principled method. Note that the signs of the estimated values for coefficients γ_{12}, γ_{22}, and γ_{32} by mode B are opposite to those by the other methods, due to five of the weights being negative for the eight indicators of ξ_2. This can be addressed by reversing the signs of the eight estimated weights.

Insert Table 6 about here

Estimated reliabilities of the four composites ($\hat{\xi}_2$, $\hat{\eta}_1$, $\hat{\eta}_2$, $\hat{\eta}_3$) by six different methods are presented in Table 6, where each method is applied to both the unstandardized (modeling $\mathbf{S}$) and standardized items (modeling $\mathbf{R}$). The six methods are respectively (1) EWC; (2) PLS$_A$ with weights estimated by PLS-SEM mode A; (3) PLS$_{A_m}$ with weights being proportional to the factor loadings $\hat{\boldsymbol{\lambda}}$ under CB-SEM presented in Table C1, where the m in the subscript is for model-implied weights (Dijkstra, 1983); (4) PLS$_B$ with weights estimated by PLS-SEM modes B; (5) PLS$_{B_A}$ with weights estimated according to section 4 of the chapter; and (6) PLS$_{B_m}$ with model-implied weights that are proportional to $\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\lambda}}$ (Dijkstra, 1983; Yuan & Deng, 2021), where $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\Psi}}$ are respectively the NML estimates of factor loadings and error variances under CB-SEM. The reliability of each composite is computed via the model-implied covariance matrix according to the NML estimates reported in Table C1. The estimated reliabilities of the individual items are also included in Table 6, which are from the default output of EQS (Bentler, 2006).

The results in Table 6 show that the reliabilities of composites by PLS$_B$ and PLS$_{B_m}$ are scale invariant while those by PLS$_{B_A}$ are close to scale invariant. The reliabilities of composites by PLS$_A$ and PLS$_{A_m}$ are rather close but neither set is scale invariant. The reliabilities of the EWCs are clearly not scale invariant. Regarding the size of the reliability estimates, those of $\hat{\xi}_2$ by both PLS$_A$ and PLS$_B$ are smaller than that of the EWC under modeling $\mathbf{S}$, and that by PLS$_B$ is also smaller than that of the EWC under modeling $\mathbf{R}$. For all the other composites, the EWCs have the smallest reliability estimates. Note that, for correctly specified models, mode B supposes to yield composites with the maximum reliabilities. The reason for PLS$_B$ to perform poorly with $\hat{\xi}_2$ is because multiple weights are negative, due to the sensitivity of the weights to model misspecification and/or sampling errors, as presented in section 6 of the chapter. For this example, all the reliability estimates for all the composites are greater than those of the individual items.

The comparison between EWC and PLS-SEM mode A in Table 6 is consistent with what was concluded by Henseler et al. (2014).

Example 6.

Mardia, Kent and Bibby (1979, Table 1.2.1) contain test scores on 5 subjects from $N = 88$ students. The five subjects are: Mechanics (y_1), Vectors (y_2), Algebra (x_1), Analysis (x_2), and Statistics (x_3). The first two scores were obtained with closed-book exams and the last three were with open-book exams. Tanaka, Watadani, and Moon (1991) fitted the dataset by a two-factor model, one factor representing the trait for taking closed-book exams, and the other representing the trait for taking open-book exams. This dataset has been used to illustrate new developments in CB-SEM and PLS-SEM (e.g., Poon & Poon, 2002; Yuan et al., 2020). We will use the dataset to compare the reliabilities of differently formulated composites. The standardized Mardia's multivariate kurtosis for this test-score dataset is $M_s = .057$, and our analysis will be conducted on the observed sample without robust transformation.

Insert Table 7 about here

Let ξ represent the trait underlying the three open-book test scores, and η represent the trait underlying the two closed-book test scores. The regression model $\eta = \gamma\xi + \zeta$ is then estimated by NML under CB-SEM, EWC regression, BFS regression, PLS-SEM modes A, $\mathrm{B_A}$ and B, respectively. The estimated regression coefficient under each method as well as the weights under the PLS methods follow the same patterns as observed with the previous organization-culture example, and we do not display them to save space. Table 7 contains the reliabilities of the 5 individual test scores estimated under CB-SEM as well as those of the weighted composites for ξ and η by 6 methods. Note that the reliability of Algebra (x_1) is .857. When working with the unstandardized variables (modeling $\mathbf{S}$), three of the six estimated reliabilities for ξ are smaller than that of the single item (x_1). Also, the reliabilities of $\hat{\xi}$ by PLS_A and PLS_{A_m} are smaller than that of the EWC. When the methods are applied to the standardized variables (modeling $\mathbf{R}$), the reliability of the EWC $\hat{\xi}$ is still smaller than that of the single item x_1, while those by the other methods are greater than that of x_1.

The results in Table 7 also show that the estimated reliabilities under PLS_{B_m} and PLS_B remain the same whether working with the standardized variables or the unstandardized variables. However, the estimated reliabilities of $\hat{\xi}$ and $\hat{\eta}$ by the other methods are not scale invariant. The causes for the reliability estimates by PLS_{B_A} not being scale invariant are sampling errors and/or possible model misspecification, and the reliabilities of EWCs and the composites under PLS_A and PLS_{A_m} are not scale invariant even under idealized conditions (correct model & without sampling error).

8. Conclusion and Discussion

In this chapter we examined several properties of PLS-SEM methodology analytically and illustrated them both numerically and via real-data examples. Although mode A with standardized variables was routinely used for models with reflective indicators, the method does not possess solid statistical or psychometric properties. Mode B possesses good psychometric and statistical properties at the level of population but it is rather sensitive to model misspecification and sampling errors. In contrast, mode B_A possesses the advantages of both modes A and B. The resulting composites by mode B_A are asymptotically most reliable, and the other results of this mode are either asymptotically scale-inverse equivariant or scale invariant and are also numerically more stable than those by mode B. We thus recommend its routine use in practice.

The sensitivity of mode B is driven by the need to maximize the relationship between predictors and the corresponding environmental variable. In particular, the weights under mode B are obtained by pulling the strength of all the indicators in each block to predict the environmental variable, which is a linear combination of the indicators of the directly connected constructs. When the model does not fit the data perfectly, mode B will pick up the additional associations not represented by the model in order to account for the relationship among the indicators. The additional associations among indicators can render the regression coefficients (the weights) negative in predicting the environment variable if the associations are not in the same direction expressed by the parameters over the existing paths. While the weights under mode A also partially maximize the associations among the indicators via the environmental variable (see e.g., Boardman et al., 1981), each indicator under mode A is solely responsible for its own weight due to being computed by simple regression, and the weight would not change its sign unless the extra association of the focal indicator with the other blocks dominates the relationship.

As with any maximization process, neither mode A nor mode B distinguishes between systematic correlations and spurious correlations due to chance errors. In order to properly utilize the maximization mechanism of PLS-SEM, additional studies are needed to separate systematic correlations from chance errors. Compared to modes A and B, mode B_A is relatively new and few studies examined its behavior with small samples or under model misspecification. Since both the systematic effect due to model misspecification and the chance effect due to sampling errors associated with $\hat{\mathbf{w}}_a$ will be inherited by $\hat{\mathbf{w}}_{b_a}$, we might expect that mode B_A will also be affected by the two types of errors, and additional studies are needed to better understand the strength of mode B_A.

There are also developments for PLS-SEM to yield estimates that are consistent with those under CB-SEM (e.g., Dijkstra & Henseler, 2015a,b; Yuan et al., 2020). As we noted in the introduction of this chapter, the values of the path coefficients under CB-SEM depend on the scales of the latent variables, and those under PLS-SEM depend on the scales of the weighted composites. Consistent estimates between the two classes of methods can be achieved by proper scaling of the latent variables or the weighted composites. Alternative methods that deviate from the standard PLS-SEM procedures can make the resulting regression equations lose the desired properties in yielding predicted values with smallest mean-squared errors (MSE), although the path coefficients are consistent with those under CB-SEM. Also, consistent path coefficients by an alternative method may not be proper if the regression equation is used for prediction or diagnosis of individuals/participants. Nevertheless, one may also develop corrections to the estimates under PLS-SEM mode B_A so that they are consistent with those defined under CB-SEM. If estimates consistent with those under CB-SEM are of primary concern, then one should start with CB-SEM rather than correcting the estimates following the PLS-SEM methodology. In addition to consistent estimates, CB-SEM also facilitates evaluations of several other features, e.g., the goodness-of-fit of the overall model structure, item reliability, unidimensionality, etc.

As noted in the introduction of this chapter, Wold (1980, 1982) recommended mode B for models with formative indicators, and there is also advice on when and how to use composites with formative indicators (e.g., Jarvis, MacKenzie & Podsakoff, 2003; Petter, Straub & Rai, 2007). Rationales for not studying formative indicators have been given in the introduction section of this chapter. Because formative indicators do not need to share a single underlying common trait, other reasons that we did not study formative indicators include: (1) The property and substantive meaning of the composites may change as the number of indicators increases, due to different compositions. (2) The meaning of a composite will also change when the structural model changes, due to the fact that weights will change when the composition of the environmental variable or the connections among the latent variables change (Treiblmaier et al., 2011). Such a dynamic nature is an integrated part of formative indicators and the corresponding composite model, which directly serves the need to maximize the relationship among different blocks of indicators. However, the data themselves do not know whether they are error-free or the variables share any common construct. The weights by mode B are still scale-inverse equivariant and the composites are scale invariant regardless of the nature of the indicators. Also, mode B may result in negative weights for truly formative indicators. Interested readers are referred to Henseler et al. (2014), Sarstedt et al. (2016); Dijkstra (2017), Hwang et al. (2020), and Cho, Sarstedt and Hwang (2022b) for detailed studies with composite models.

We have showed statistical and psychometric properties of PLS-SEM analytically and numerically as well as by real data examples. While PLS-SEM mode B enjoys many theoretical and/or asymptotic advantages, results also showed that the method is rather sensitive to sampling errors. Further analytical or Monte Carlo studies are needed to see the speed for mode B to converge to its asymptotic results or when the method yields similar empirical results as BFS regression.

References

Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411–423. doi: 10.1037/0033-2909.103.3.411

Bentler, P.M. (1968). Alpha-maximized factor analysis (Alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika, 33*(3), 335–345. doi: 10.1007/BF02289328

Bentler, P.M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bergami, M., & Bagozzi, R.P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization. *British Journal of Social Psychology, 39*(4), 555–577. doi: 10.1348/014466600164633

Boardman, A.E., Hui, B.S., & Wold, H. (1981). The partial least squares-fix point method of estimating interdependent systems with latent variables. *Communications in Statistics-Theory and Methods, 10*(7), 613–639. doi: 10.1080/03610928108828062

Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G.A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates.

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika, 47*(1), 243–272. doi: 10.1007/s41237-019-00098-0

Cho, G., Kim, S., Lee, J., Hwang, H., Sarstedt, M., & Ringle, C.M. (2022a). A comparative study of the predictive power of component-based approaches to structural equation modeling. *European Journal of Marketing.* Advance online publication.

Cho, G., Sarstedt, M., & Hwang, H. (2022b). A comparative evaluation of factor- and component-based structural equation modelling approaches under (in)correct construct representations. *British Journal of Mathematical and Statistical Psychology, 75*(2), 220–251. doi: 10.1111/bmsp.12255

Cochran, W.G. (1970). Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association, 65*(329), 22–34. doi: 10.1080/01621459.1970.10481059

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement, 76*, 741–770. doi:10.1177/0013164415607618

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares estimates. *Journal of Econometrics, 22*, 67–90. doi: 10.1016/0304-4076(83)90094-5

Dijkstra, T.K. (2017). A perfect match between a model and a mode. In H. Latan & R. Noonan (Eds.), *Partial least squares path modeling: Basic concepts, methodological issues and applications* (pp. 55–80). Cham, Switzerland: Springer.

Dijkstra, T.K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics and Data Analysis, 81*, 10–23. doi: 10.1016/j.csda.2014.07.008

Dijkstra, T.K., & Henseler, J. (2015b). Consistent partial least squares path modeling. *MIS Quarterly, 39*(2), 297–316. doi: 10.25300/MISQ/2015/39.2.02

Fuller, W.A. (1987). *Measurement error models*. New York: Wiley.

Hair, J.F., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed). Thousand Oaks, CA: Sage. ISBN 9781483377445

Henseler, J. (2021). *Composite-based structural equation modeling*. New York: Guilford. ISBN: 9781462545605

Henseler, J., Dijkstra, T.K., Sarstedt, M., Ringle, C.M., Diamantopoulos, A., Straub, D.W., Ketchen, D.J., Hair, J.F., Hult, G.T.M., & Calantone, R.J. (2014). Common beliefs and reality about partial least squares: Comments on Rönkkö & Evermann (2013). *Organizational Research Methods, 17*(2), 182–209. doi: 10.1177/1094428114526928

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi: 10.1080/10705519909540118

Huber, P.J. (1981). *Robust statistics*. New York: Wiley.

Hwang, H., Sarstedt, M., Cheah, J.H., & Ringle, C.M. (2020). A concept analysis of methodological research on composite-based structural equation modeling: Bridging PLSPM and GSCA. *Behaviormetrika, 47*, 219–241. doi: 10.1007/s41237-019-00085-5

Jarvis, C.B., MacKenzie, S.B., & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199–218. doi: 10.1086/376806

Loehlin, J.C., & Beaujean, A.A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). New York: Routledge.

Mardia, K.V. (1970). Measure of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530. doi:10.1093/biomet/57.3.519

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate analysis*. New York: Academic Press.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38*(1), 113–139. doi: 10.1207/S15327906MBR3801_5

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623–656. doi: 10.2307/25148814

Poon, W.-Y., & Poon, Y.S. (2002). Influential observations in the estimation of mean vector and covariance matrix. *British Journal of Mathematical and Statistical Psychology, 55*, 177–192. doi: 10.1348/000711002159644

Rönkkö, M., Lee, N., Evermann, J., McIntosh, C. Antonakis, J. (2022). Marketing or methodology? exposing the fallacies of PLS with simple demonstrations. *European Journal of Marketing.*

Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O., & Gudergan, S.P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010. doi: 10.1016/j.jbusres.2016.06.007

Schneeweiss, H. (1993). Consistency at large in models with latent variables. In K. Haagen, D. J. Bartholomew, & M. Deistler (Eds.), *Statistical modelling and latent variables* (pp. 299–320). Amsterdam: Elsevier Science Publishers.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*, 563–575. doi:10.1007/BF02296196

Tanaka, Y., Watadani, S., & Moon, S.H. (1991). Influence in covariance structure analysis: With an application to confirmatory factor analysis. *Communications in Statistics: Theory and Method, 20*, 3805–3821. doi: 10.1080/03610929108830742

Treiblmaier, H., Bentler, P.M., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling, 18*(1), 1–17. doi: 10.1080/10705511.2011.532693

Vinzi, V.E., Trinchera, L., & Amato, S. (2010). PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. In V.E. Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications* (pp. 47–82). Berlin: Springer.

Wold, H. (1980). Model construction and evaluation when theoretical knowledge is scarce. In J. Kmenta & J.B. Ramsey (Eds.), *Evaluation of econometric models* (pp. 47–74). New York: Academic Press.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation, Part II* (pp. 1–54). Amsterdam: North-Holland.

Yuan, K.-H., & Bentler, P.M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*, 251–259. doi: 10.1007/BF02294845

Yuan, K.-H., Chan, W., & Bentler, P.M. (2000). Robust transformation with applications to structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 53*, 31–50. doi: 10.1348/000711000159169

Yuan, K.-H., & Deng, L. (2021). Equivalence of partial-least-squares SEM and the methods of factor-score regression. *Structural Equation Modeling, 28*(4), 557–571. doi: 10.1080/10705511.2021.1894940

Yuan, K.-H., & Fang, Y. (2022). Which method delivers greater signal-to-noise ratio: Structural equation modeling or regression analysis with weighted composites? *British Journal of Mathematical and Statistical Psychology.* doi: 10.1111/bmsp.12293

Yuan, K.-H., & Gomer, B. (2021). An overview of applied robust methods. *British Journal of Mathematical and Statistical Psychology, 74*, 199–246. doi: 10.1111/bmsp.12230

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). Assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology, 33*, 241–265. doi: 10.1111/j.0081-1750.2003.00132.x

Yuan, K.-H., Wen, Y., & Tang, J. (2020). Regression analysis with latent variables by partial least squares and four other composite scores: Consistency, bias and correction. *Structural Equation Modeling, 27*(3), 333–350. doi: 10.1080/10705511.2019.1647107

Yuan, K.-H., Wen, Y., & Tang, J. (2022). Sensitivity analysis of the weights of the composites under partial least-squares approach to structural equation modeling. *Structural Equation Modeling*. doi: 10.1080/10705511.2022.2106487

Appendix A. Population and sample covariance matrices

The tables in this appendix contain the population and sample covariance matrices that are used to compute the results in Tables 2 and 3, respectively. The values for the population covariance matrix (Table A1) are exact, and those for the sample covariance matrix are rounded. Note that the sample covariance matrix is unbiased.

Table A1. The population covariance matrix $\mathbf{\Sigma}$

	x_1	x_2	x_3	y_1	y_2	y_3	y_4
x_1	1.1900	0.8000	0.9600	0.5600	0.6720	0.4480	0.8400
x_2	0.8000	1.6000	1.2000	0.7000	0.8400	0.5600	1.0500
x_3	0.9600	1.2000	1.8400	0.8400	1.0080	0.6720	1.2600
y_1	0.5600	0.7000	0.8400	1.4700	1.0680	0.7120	1.3350
y_2	0.6720	0.8400	1.0080	1.0680	1.9316	0.8544	1.6020
y_3	0.4480	0.5600	0.6720	0.7120	0.8544	1.1696	1.0680
y_4	0.8400	1.0500	1.2600	1.3350	1.6020	1.0680	2.4025

Table A2. The (unbiased) sample covariance matrix $\mathbf{S}$

	x_1	x_2	x_3	y_1	y_2	y_3	y_4
x_1	1.5429	1.1706	1.4051	0.4285	0.7840	0.5716	1.1351
x_2	1.1706	1.6088	1.4303	0.4590	0.8221	0.7012	1.1209
x_3	1.4051	1.4303	2.0897	0.5471	1.0944	0.7674	1.3803
y_1	0.4285	0.4590	0.5471	1.4250	0.8699	0.6080	1.1202
y_2	0.7840	0.8221	1.0944	0.8699	1.7641	0.8380	1.4730
y_3	0.5716	0.7012	0.7674	0.6080	0.8380	1.5575	1.0482
y_4	1.1351	1.1209	1.3803	1.1202	1.4730	1.0482	2.3187

Appendix B. The SAS IML code that generates the sample covariance matrix in Table A2

```
proc iml;
n=100; *sample size;
p_x=3;
p_y=4;
p=p_x+p_y; *number of variables;
*-------------------------;
*population values;
lamb_x0={0.8, 1.0, 1.2}; *loadings;
lamb_y0={1.0, 1.2, 0.8, 1.5};
lamb0=(lamb_x0||j(3,1,0))//(j(4,1,0)||lamb_y0);
phi_11=1.0; *factor variance;
gamma=.7; *regression coefficient;
sig2_zeta=.40; *prediction error variance;

phi_12=gamma;
phi_21=phi_12;
phi_22=gamma*gamma+sig2_zeta;

phi0=
(phi_11||phi_12)//
(phi_21||phi_22);
psi0={.55,.6,.4,.58,.65,.6,.4}; *measurement error variances;
psi_mat0=diag(psi0);

Sig_0=lamb0*Phi0*(lamb0`)+psi_mat0;*population covariance matrix;

call eigen(sval0,svec0,Sig_0); *eigenvalue decomposition;
sig_012=svec0*diag(sqrt(sval0))*svec0`; *Sig_0^{1/2};
*-------------------------;
seed=1111111111;
z=normal(j(n,p,seed));
*a 100 by 7 matrix of independent random numbers following N(0,1);
x=z*sig_012; *the observed sample;
*each row of x follows a 7-variate normal distribution N(0,Sig0);
print x;
scov_x=x`*(i(n)-j(n,n,1)/n)*x/(n-1); *the sample covariance matrix;
print scov_x;
```

Appendix C. Normal-distribution-based maximum likelihood (NML) estimates for the model in Figure 3

The table in this appendix contains the NML estimates by fitting the model in Figure 3 to the robustly transformed sample via H(.25). The dataset was originally presented by Bergami and Bagozzi (2000). The data and model used in this chapter are from the book by Henseler (2021). The likelihood ratio statistic and goodness-of-fit indices (RMSEA and CFI) were reported in the chapter and are also included in the table.

Table C1. Parameter estimates (est), their SEs (se) and z-statistics for the CB-SEM model represented by Figure 3 ($p = 21$, $N = 305$, $T_{ml} = 613.756$, $df = 183$, p-value=0; RMSEA=.088, and CFI=.871). The results are obtained using NML with a robustly transformed sample via H(.25).

θ	est	se	z	θ	est	se	z
$\lambda_{x1,1}$	1.000			ψ_{x1}	0		
$\lambda_{x2,2}$	1.000			ψ_{x2}	0.217	0.019	11.341
$\lambda_{x3,2}$	1.121	0.086	13.105	ψ_{x3}	0.197	0.018	10.960
$\lambda_{x4,2}$	1.119	0.093	12.082	ψ_{x4}	0.281	0.025	11.373
$\lambda_{x5,2}$	1.036	0.075	13.862	ψ_{x5}	0.125	0.012	10.475
$\lambda_{x6,2}$	1.132	0.083	13.689	ψ_{x6}	0.160	0.015	10.605
$\lambda_{x7,2}$	1.266	0.090	14.102	ψ_{x7}	0.168	0.016	10.268
$\lambda_{x8,2}$	1.210	0.096	12.597	ψ_{x8}	0.275	0.025	11.189
$\lambda_{x9,2}$	1.107	0.078	14.186	ψ_{x9}	0.123	0.012	10.187
$\lambda_{y1,1}$	1.000			ψ_{y1}	0.285	0.026	10.940
$\lambda_{y2,1}$	0.922	0.089	10.379	ψ_{y2}	0.313	0.028	11.257
$\lambda_{y3,1}$	0.739	0.087	8.480	ψ_{y3}	0.383	0.033	11.771
$\lambda_{y4,1}$	1.347	0.103	13.112	ψ_{y4}	0.203	0.023	8.756
$\lambda_{y5,1}$	1.340	0.105	12.707	ψ_{y5}	0.254	0.027	9.494
$\lambda_{y6,1}$	0.964	0.089	10.820	ψ_{y6}	0.292	0.026	11.071
$\lambda_{y7,2}$	1.000			ψ_{y7}	0.378	0.036	10.513
$\lambda_{y8,2}$	0.980	0.098	10.029	ψ_{y8}	0.282	0.028	9.975
$\lambda_{y9,2}$	1.118	0.103	10.814	ψ_{y9}	0.225	0.027	8.450
$\lambda_{y10,2}$	0.828	0.094	8.844	ψ_{y10}	0.347	0.032	10.981
$\lambda_{y11,3}$	1.000			ψ_{y11}	0.325	0.036	8.960
$\lambda_{y12,3}$	1.183	0.151	7.811	ψ_{y12}	0.290	0.042	6.908
ϕ_{11}	0.251	0.020	12.329				
ϕ_{22}	0.231	0.033	7.068				
γ_{21}	-0.182	0.052	-3.498	β_{21}	0.713	0.086	8.309
γ_{31}	0.135	0.057	2.365	β_{31}	-0.564	0.084	-6.735
γ_{12}	0.372	0.072	5.206	$\sigma^2_{\zeta_1}$	0.233	0.036	6.567
γ_{22}	0.222	0.063	3.551	$\sigma^2_{\zeta_2}$	0.103	0.021	4.855
γ_{32}	-0.197	0.068	-2.881	$\sigma^2_{\zeta_3}$	0.101	0.026	3.933

Table 4. Estimated weights of composites by three methods for the model in Figure 3 ($p = 21$, $N = 305$). The results are obtained with a robustly transformed sample via H(.25).

Indicator	Modeling $\mathbf{S}$ ($\hat{\mathbf{w}}_s$)			Modeling $\mathbf{R}$ ($\hat{\mathbf{w}}_r$)			$\mathbf{D}_s\hat{\mathbf{w}}_s$		
	$\hat{w}_a$	$\hat{w}_{b_a}$	$\hat{w}_b$	$\hat{w}_a$	$\hat{w}_{b_a}$	$\hat{w}_b$	$\hat{w}_a$	$\hat{w}_{b_a}$	$\hat{w}_b$
$x_{1,1}$	2.001	2.001	2.001	1.000	1.000	1.000	1.000	1.000	1.000
$x_{2,2}$	0.166	0.101	0.140	0.122	0.069	0.094	0.111	0.068	0.094
$x_{3,2}$	0.196	0.131	0.273	0.137	0.091	0.190	0.136	0.091	0.190
$x_{4,2}$	0.251	0.222	-0.589	0.161	0.163	-0.443	0.189	0.167	-0.443
$x_{5,2}$	0.199	0.251	-0.182	0.159	0.155	-0.111	0.121	0.153	-0.111
$x_{6,2}$	0.221	0.230	-0.480	0.160	0.160	-0.324	0.149	0.155	-0.324
$x_{7,2}$	0.262	0.350	-0.467	0.174	0.258	-0.342	0.192	0.256	-0.342
$x_{8,2}$	0.251	0.181	0.003	0.155	0.137	0.003	0.196	0.141	0.003
$x_{9,2}$	0.217	0.294	-0.350	0.165	0.184	-0.222	0.138	0.187	-0.222
$y_{1,1}$	0.257	0.231	0.179	0.211	0.171	0.132	0.190	0.171	0.132
$y_{2,1}$	0.257	0.241	0.289	0.215	0.181	0.212	0.188	0.177	0.212
$y_{3,1}$	0.209	0.154	0.196	0.176	0.113	0.142	0.152	0.112	0.142
$y_{4,1}$	0.350	0.448	0.445	0.257	0.368	0.368	0.289	0.370	0.368
$y_{5,1}$	0.345	0.342	0.369	0.245	0.289	0.315	0.294	0.292	0.315
$y_{6,1}$	0.236	0.194	0.132	0.196	0.142	0.097	0.173	0.142	0.097
$y_{7,2}$	0.428	0.330	0.299	0.305	0.265	0.242	0.347	0.268	0.242
$y_{8,2}$	0.417	0.428	0.415	0.326	0.314	0.309	0.310	0.318	0.309
$y_{9,2}$	0.474	0.597	0.659	0.363	0.453	0.501	0.360	0.453	0.501
$y_{10,2}$	0.363	0.314	0.282	0.292	0.238	0.208	0.267	0.230	0.208
$y_{11,3}$	0.678	0.574	0.601	0.522	0.412	0.441	0.498	0.422	0.441
$y_{12,3}$	0.871	0.956	0.935	0.649	0.743	0.720	0.670	0.735	0.720

Note: $\mathbf{S} = (s_{ij})$ and $\mathbf{R} = (r_{ij})$ are respectively the sample covariance and correlation matrices of the robustly transformed sample, and the d_s is the square root of the corresponding diagonal element of $\mathbf{S}$.

Table 5. Estimated path coefficients and the corresponding R-squares for the model in Figure 3 ($p = 21,\ N = 305$). The results are obtained with a robustly transformed sample via H(.25).

	Modeling $\mathbf{S}$				Modeling $\mathbf{R}$			
parameter	EWC$_{reg}$	PLS$_A$	PLS$_{B_A}$	PLS$_B$	EWC$_{reg}$	PLS$_A$	PLS$_{B_A}$	PLS$_B$
γ_{21}	-0.169	-0.145	-0.151	-0.148	-0.110	-0.143	-0.151	-0.148
γ_{31}	0.127	0.114	0.117	0.112	0.084	0.111	0.117	0.112
γ_{12}	0.365	0.327	0.325	-0.337	0.327	0.326	0.325	-0.337
γ_{22}	0.258	0.208	0.200	-0.191	0.236	0.210	0.200	-0.191
γ_{32}	-0.206	-0.178	-0.185	0.191	-0.190	-0.179	-0.185	0.191
β_{21}	0.572	0.540	0.554	0.556	0.576	0.538	0.554	0.556
β_{31}	-0.442	-0.430	-0.431	-0.430	-0.447	-0.423	-0.430	-0.430
$R^2_{\eta_1}$	0.107	0.107	0.106	0.114	0.106	0.107	0.106	0.114
$R^2_{\eta_2}$	0.417	0.425	0.437	0.437	0.415	0.424	0.437	0.437
$R^2_{\eta_3}$	0.264	0.277	0.282	0.287	0.259	0.270	0.281	0.287

Note: $\mathbf{S} = (s_{ij})$ and $\mathbf{R} = (r_{ij})$ are respectively the sample covariance and correlation matrices of the robustly transformed sample.

Table 6. Reliabilities of individual items (ρ_j) as well as of composites (ξ_2, η_2, η_2, η_3) by six methods for the model in Figure 3 ($p = 21$, $N = 305$). The results are obtained with a robustly transformed sample via H(.25).

Item	ρ_j	Item	ρ_j	Item	ρ_j	Item	ρ_j
$x_{2,2}$	0.515	$y_{1,1}$	0.482	$y_{7,2}$	0.430	$y_{11,3}$	0.402
$x_{3,2}$	0.595	$y_{2,1}$	0.419	$y_{8,2}$	0.492	$y_{12,3}$	0.512
$x_{4,2}$	0.507	$y_{3,1}$	0.274	$y_{9,2}$	0.613		
$x_{5,2}$	0.665	$y_{4,1}$	0.704	$y_{10,2}$	0.360		
$x_{6,2}$	0.649	$y_{5,1}$	0.653				
$x_{7,2}$	0.688	$y_{6,1}$	0.458				
$x_{8,2}$	0.551						
$x_{9,2}$	0.696						

	Modeling **S**				Modeling **R**			
Method	$\hat{\xi}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	$\hat{\xi}_2$	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$
EWC	0.923	0.859	0.781	0.628	0.925	0.854	0.781	0.627
PLS_{A_m}	0.923	0.873	0.787	0.632	0.927	0.867	0.789	0.630
PLS_{A}	0.921	0.872	0.787	0.633	0.927	0.865	0.788	0.632
PLS_{B_m}	0.929	0.877	0.795	0.633	0.929	0.877	0.795	0.633
PLS_{B}	0.803	0.870	0.794	0.631	0.803	0.870	0.794	0.631
PLS_{B_A}	0.924	0.876	0.795	0.630	0.925	0.876	0.794	0.629

Note: $\mathbf{S} = (s_{ij})$ and $\mathbf{R} = (r_{ij})$ are respectively the sample covariance and correlation matrices of the robustly transformed sample.

Table 7. Reliabilities of individual items (ρ_j) as well as of composites (ξ, η) by six methods for test-score data ($p = 5$, $N = 88$).

Individual item			Modeling $\mathbf{S}$		Modeling $\mathbf{R}$	
Item	ρ_j	Method	$\hat{\xi}$	$\hat{\eta}$	$\hat{\xi}$	$\hat{\eta}$
x_1	0.857	EWC	0.823	0.699	0.852	0.715
x_2	0.599	PLS_{A_m}	0.808	0.687	0.865	0.719
x_3	0.526	PLS_A	0.814	0.687	0.870	0.720
y_1	0.491	PLS_{B_m}	0.896	0.724	0.896	0.724
y_2	0.624	PLS_B	0.884	0.724	0.884	0.724
		PLS_{B_A}	0.858	0.724	0.876	0.724

Note: $\mathbf{S} = (s_{ij})$ and $\mathbf{R} = (r_{ij})$ are respectively the sample covariance and correlation matrices of the observed sample.