

Consistent and asymptotically normal PLS estimators for linear structural equations.

Theo K. Dijkstra¹ and Jörg Henseler²

¹University of Groningen, NL, Faculty of Economics and Business

²Radboud University Nijmegen, NL, Institute for Management Research

June 17, 2012

This paper introduces a vital extension to partial least squares path modeling: consistency. While maintaining all strengths of PLS, the consistent version provides two key improvements. Path coefficients, parameters of simultaneous equations, construct correlations, and indicator loadings are estimated consistently. And the global goodness of fit of the structural model can now be assessed, which makes PLS suitable for confirmatory research. A Monte Carlo simulation illustrates the new approach and compares it with covariance-based structural equation modeling.

Keywords: PLS; consistent partial least squares; SEM; variance-based structural equation modeling; Monte Carlo simulation

Consistent Partial Least Squares

INTRODUCTION

Structural equation modeling (SEM) has become *the* tool of the trade in survey-based research. Researchers embrace its abilities, such as modeling latent variables, correcting for measurement errors, and simultaneously estimating parameters of entire theories. Two families of structural equation modeling techniques prevail (Chin, 1998; Reinartz et al., 2009): covariance-based SEM and variance-based SEM. Variance-based SEM appears to be experiencing increasing popularity, as evidenced by recent methodological advances (c.f. Hwang et al., 2010; Lu et al., 2011) and frequent application (Ringle et al., 2012; Hair et al., 2012). Researchers appreciate the advantages of variance-based SEM, such as the lack of convergence problems and factor indeterminacy (Henseler, 2010), relatively mild distributional assumptions (Reinartz et al., 2009), and the possibility of estimating models having more variables or parameters than observations. Variance-based SEM includes many different techniques, such as regression based on sum scores or principal components (Tenenhaus, 2008), partial least squares path modeling (PLS, see Wold, 1982), and generalized structured component analysis (Hwang et al., 2010; Henseler, 2012). All variance-based SEM techniques have in common the characteristic of approximating latent variables using linear composites of observed variables.

Among variance-based SEM techniques, PLS is regarded as the “most fully developed and general system” (McDonald, 1996, p. 240). MIS research relies strongly on PLS as a method of analysis (Marcoulides and Saunders, 2006; Ringle et al., 2012), and many extensions and advances of PLS path modeling can be credited to MIS researchers. Concretely, PLS-based approaches for multigroup analysis (Chin and Dibbern, 2010; Keil et al., 2000; Qureshi and Compeau, 2009), testing of moderating effects (Chin et al., 2003; Goodhue et al., 2007; Henseler and Chin, 2010), assessing common method bias (Liang et al., 2007; Chin et al., forthcoming), testing for measurement invariance (Hsieh

et al., 2008), modeling nonlinear relationships (Henseler et al., 2012), and analysis of hierarchical component models (Ringle et al., 2012; Wetzels et al., 2009) were developed, discussed, or improved within the MIS domain.

Yet the use of PLS as an estimator for structural equation models is not without severe disadvantages. First, PLS estimates – in particular, path coefficients and loadings – are not consistent, but are only consistent at large (Wold, 1982). Consequently, “[p]arameter estimates for paths between observed variables and latent variable proxies are biased upward in PLS (away from zero), while parameter estimates for paths between proxies are attenuated” (Gefen et al., 2011, p. vi). Second, PLS does not provide overall goodness-of-fit measures, meaning that empirically testing or comparing theories, as is done with covariance-based SEM, is not possible (Fornell and Bookstein, 1982; Henseler and Sarstedt, 2012). Out of these two deficiencies of PLS, lack of consistency is probably the most severe, because of its adverse consequences for substantial research findings. If PLS underestimates the true parameter, Type-II errors are likely. If PLS overestimates the true parameter, the Type-I error is inflated. Finally, the lack of consistency entails that meta studies based on PLS estimates are not guaranteed to come closer to the true value than single studies.

This paper introduces an important advancement to PLS that overcomes the aforementioned deficiencies: consistency. While maintaining all of the strengths of PLS, consistent PLS provides several key improvements. It consistently estimates the path coefficients, construct correlations, and indicator loadings; it allows non-recursive models to be estimated; and it provides a global assessment of goodness of fit. A Monte Carlo simulation sheds light on the performance of the consistent version PLSc relative to covariance-based SEM.

CONSISTENT PARTIAL LEAST SQUARES

Herman Wold developed partial least squares in the sixties, seventies, and eighties of the previous century to analyze high-dimensional data reflecting unobserved entities that are interconnected in path diagrams (Wold, 1966, 1975, 1982). Principal components and canonical variables were the main source of inspiration. Linear compounds are constructed to serve as proxies or stand-ins for the latent variables, leading to straightforward estimates of structural parameters such as path coefficients and loadings. Compound weights are generated using a variety of alternating least squares algorithms. They are cycles of regressions that update natural subsets of weights in turn, coming to a halt when consecutive estimates no longer change significantly. Convergence, which is the rule, is typically very fast. PLS has become a vibrant field of both applied and theoretical research; see, for example, Tenenhaus et al. (2005) and the Handbook of Partial Least Squares by Vinzi et al. (2010) for overviews.

In the spirit of principal components and canonical variables, PLS has been very useful as a method to extract information from high-dimensional data. However, as a means to estimating the parameters of latent variable models, PLS leaves something to be desired: the relationships between linear compounds can never duplicate the relationships between the latent variables, except for sets of measure zero in the parameter space. In fact, in linear factor models, PLS tends to overestimate the absolute value of loadings and underestimate the multiple and bivariate (absolute) correlation coefficients. Dijkstra (1981, 1983, 2010) shows how to correct for this tendency. The consistent version of PLS is denoted by PLS_c. Subsequent sections outline the PLS_c approach and show that it gives consistent and asymptotically normal estimators (CAN-estimators) for the parameters of interest in this paper.

Weight vectors

A starting point for PLS analysis is the so-called “basic design,” in essence a second-order factor model. A number of i.i.d. vectors of indicators are assumed to exist from a population with finite moments of at least order two (the precise order depending on other distributional assumptions or requirements). All indicators have zero mean and unit variance. The vector of indicators y is composed of at least two subvectors, each measuring a unique latent variable, and each subvector contains at least two components. For the i^{th} subvector y_i we have:

$$y_i = \lambda_i \cdot \eta_i + \epsilon_i \tag{1}$$

where the “loading vector” λ_i and the “vector of idiosyncratic errors” ϵ_i have the same dimensions as y_i , and the unobservable “latent variable” η_i is real-valued. For convenience, the sufficient but by no means necessary assumption is made that all components of all error vectors are mutually independent and are independent of all latent variables. The latter have zero mean and unit variance. The correlations between η_i and η_j are denoted by ρ_{ij} . They are collected in a matrix Φ , $\Phi := (\rho_{ij})$. At this moment, the nature of the relationships between the latent variables – whether linear or nonlinear – is not of concern.

A particular set of easy implications is that the covariance matrix Σ_{ii} of y_i can be written as:

$$\Sigma_{ii} := E y_i y_i^\top = \lambda_i \lambda_i^\top + \Theta_i \tag{2}$$

where Θ_i is diagonal with non-negative diagonal elements, and we have for the covariance between y_i and y_j :

$$\Sigma_{ij} := E y_i y_j^\top = \rho_{ij} \lambda_i \lambda_j^\top. \tag{3}$$

The sample counterparts of Σ_{ii} and Σ_{ij} are denoted by S_{ii} and S_{ij} , respectively. Without repeatedly saying so, the sample data are assumed to be standardized before being analyzed. Therefore, the observed data have zero mean and unit (sample) variance. Note

that the assumptions made so far entail that the sample counterparts are consistent and asymptotically normal estimators of the theoretical variance and covariance matrices.

PLS features a number of iterative *fixed-point* algorithms, of which the so-called “mode A algorithm” is selected. In general, the “mode A algorithm” is numerically the most stable algorithm (for discussions of PLS’ other modes, see Lohmöller, 1989). As a rule, the algorithm converges and is usually very fast (for example, for the models analyzed in this paper, no more than five iterations are needed to obtain five significant decimals). The outcome is an estimated weight vector \hat{w} with typical subvector \hat{w}_i of the same dimensions as y_i . With these weights, *sample proxies* are defined for the latent variables: $\hat{\eta}_i := \hat{w}_i^\top y_i$ for η_i , with the customary normalization of a unit sampling variance, so $\hat{w}_i^\top S_{ii} \hat{w}_i = 1$. In Wold’s PLS approach, the $\hat{\eta}_i$ ’s replace the unobserved latent variables, and the loadings and structural parameters are estimated “directly,” in contrast to covariance-based SEM, which, for example, follows the opposite order. In “mode A,” for each i :

$$\hat{w}_i \propto \sum_{j \in C(i)} \text{sgn}_{ij} \cdot S_{ij} \hat{w}_j. \quad (4)$$

Here, sgn_{ij} is the sign of the sample correlation between $\hat{\eta}_i$ and $\hat{\eta}_j$ and $C(i)$ is a set of indices of latent variables. Traditionally, $C(i)$ contains the indices of latent variables “adjacent” to η_i , in other words, those that appear on the *other* side of the structural or path equations in which η_i appears. This setup is not always a good idea, particularly when the correlations between the indicators of η_i and the indicators of its adjacent variables are weak. In general, the use of *all* $j \neq i$ is suggested in Equation (4). Clearly, \hat{w}_i is obtained by a regression of the indicators y_i on the “sign-weighted sum” of the selected proxies: $\sum_{j \in C(i)} \text{sgn}_{ij} \cdot \hat{\eta}_j$. Other versions exist (for example, with correlation weights); this version, which is the original, is one of the simplest (see Wold, 1982). Little motivation exists in the PLS literature for the coefficients of $S_{ij} \hat{w}_j$. The particular choice can be shown to be irrelevant for the probability limits of the estimators. The algorithm takes an arbitrary starting vector and then basically follows the sequence of regressions

for each i , each time inserting updates when available (or after each full round; the precise implementation is not important).

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

Factor loadings

Given the consistency, S may be replaced by Σ to obtain:

$$\bar{w}_i := \text{plim } \hat{w}_i = \frac{\lambda_i}{\sqrt{\lambda_i^\top \Sigma_{ii} \lambda_i}}. \quad (5)$$

Note that the probability limits of the weight vectors are proportional to the loadings. In PLS, the loadings are estimated by a regression of the indicators y_i on their direct sample proxy $\hat{\eta}_i$; however, because doing so in general removes the proportionality, *this tradition is not followed (for mode A)*.

As in Dijkstra (1981, 2010), the following definition is proposed:

$$\hat{\lambda}_i := \hat{c}_i \cdot \hat{w}_i, \quad (6)$$

where the scalar \hat{c}_i is such that the off-diagonal elements of S_{ii} are reproduced as best as possible in a least squares sense. Therefore, the Euclidean distance is minimized between:

$$\left[S_{ii} - \text{diag}(S_{ii}) \right] \quad \text{and} \quad \left[(c_i \cdot \hat{w}_i)(c_i \cdot \hat{w}_i)^\top - \text{diag}\left((c_i \cdot \hat{w}_i)(c_i \cdot \hat{w}_i)^\top\right) \right]$$

as a function of c_i and the following is obtained:

$$\hat{c}_i := \sqrt{\frac{\hat{w}_i^\top (S_{ii} - \text{diag}(S_{ii})) \hat{w}_i}{\hat{w}_i^\top (\hat{w}_i \hat{w}_i^\top - \text{diag}(\hat{w}_i \hat{w}_i^\top)) \hat{w}_i}}. \quad (7)$$

In sufficiently large samples, \widehat{c}_i will be well-defined, real, and positive. (In all samples in this paper and those in another study, \widehat{c}_i attained proper values.) Its calculation does not require additional numerical optimization. Verifying that the correction does its job is straightforward by replacing S_{ii} by Σ_{ii} and \widehat{w}_i by \bar{w}_i : the matrix in the denominator equals the matrix in the numerator, apart from a factor $(\lambda_i^\top \Sigma_{ii} \lambda_i)^{-1}$; therefore:

$$\bar{c}_i := \text{plim } \widehat{c}_i = \sqrt{\lambda_i^\top \Sigma_{ii} \lambda_i}. \quad (8)$$

Now, in particular:

$$\text{plim } \widehat{\lambda}_i = \text{plim } (\widehat{c}_i \cdot \widehat{w}_i) = \bar{c}_i \cdot \bar{w}_i = \lambda_i. \quad (9)$$

Correlations between latent variables

Defining a “population proxy” $\bar{\eta}_i$ by $\bar{\eta}_i := \bar{w}_i^\top y_i$ is useful. Clearly, the squared correlation between a population proxy and its corresponding latent variable is:

$$R^2(\eta_i, \bar{\eta}_i) = \left(\bar{w}_i^\top \lambda_i \right)^2, \quad (10)$$

which equals:

$$\frac{(\lambda_i^\top \lambda_i)^2}{\lambda_i^\top \Sigma_{ii} \lambda_i} = \frac{(\lambda_i^\top \lambda_i)^2}{(\lambda_i^\top \lambda_i)^2 + \lambda_i^\top \Theta_i \lambda_i}. \quad (11)$$

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

$$R^2(\bar{\eta}_i, \bar{\eta}_j) = \left(\bar{w}_i^\top \Sigma_{ij} \bar{w}_j \right)^2 = \rho_{ij}^2 \cdot R^2(\eta_i, \bar{\eta}_i) \cdot R^2(\eta_j, \bar{\eta}_j) \quad (12)$$

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

$$R^2(\eta_i, \bar{\eta}_i) = \left(\bar{w}_i^\top \lambda_i \right)^2 = \left(\bar{w}_i^\top \cdot (\bar{w}_i \cdot \bar{c}_i) \right)^2 = \left(\bar{w}_i^\top \bar{w}_i \right)^2 \cdot \bar{c}_i^2 \quad (13)$$

enabling an estimation of the (squared) quality of the proxies consistently by:

$$R^2(\widehat{\eta}_i, \overline{\eta}_i) := \left(\widehat{w}_i^\top \widehat{w}_i\right)^2 \cdot \widehat{c}_i^2. \quad (14)$$

Moreover, with:

$$R^2(\widehat{\eta}_i, \overline{\eta}_j) := \left(\widehat{w}_i^\top S_{ij} \widehat{w}_j\right)^2, \quad (15)$$

see Equation (12); the correlations between the latent variables can be consistently estimated using:

$$\widehat{\rho}_{ij}^2 := \frac{R^2(\widehat{\eta}_i, \overline{\eta}_j)}{R^2(\widehat{\eta}_i, \overline{\eta}_i) \cdot R^2(\overline{\eta}_j, \overline{\eta}_j)}. \quad (16)$$

Therefore, $E\eta_i\eta_j$ is estimated using the sample covariance between the proxies $\widehat{\eta}_i$ and $\widehat{\eta}_j$, each divided by its estimated quality. Finally, let $\widehat{\Phi} := (\widehat{\rho}_{ij})$.

Note that standard PLS software for mode A produces all of the necessary ingredients for a consistent estimation; all that is required is a simple rescaling of the weight vectors.

Simultaneous equation systems

It is important to note that when the latent variables are mutually related through linear equations, whether recursively or with feedback patterns, CAN-estimates of their coefficients are also obtainable, provided that they are identifiable from the second-order moment matrix of the latent variables through smooth (locally continuously differentiable) mappings. In principle, partially identifiable structures can also be handled using Bekker and Dijkstra (1990) and Bekker et al. (1994).

The method of choice in this study is the old econometric workhorse 2SLS, two-stage least squares. 2SLS estimates each equation separately, is a limited-information technique, is probably the simplest estimation method around, and does not – despite appearances – require additional iterations. Boardman et al. (1981) employed an iterative version of 2SLS, Wold’s fix-point method, using the original PLS input uncorrected for inconsistency.

For completeness' sake, the 2SLS method is specified here. To this end, consider the “linear structural equations”:

$$\eta_{p+1:p+q} = B \cdot \eta_{p+1:p+q} + \Gamma \cdot \eta_{1:p} + \zeta, \quad (17)$$

where η is partitioned into a vector of p components, $\eta_{1:p}$, the exogenous latent variables, and a vector of q components, $\eta_{p+1:p+q}$, the endogenous latent variables. The residual vector ζ has zero mean and is uncorrelated with (or independent of) $\eta_{1:p}$. The $q \times q$ matrix B captures the feedback or reciprocal relationships between the endogenous variables, and is such that the inverse of $I - B$ exists. The latter assumption enables one to write, with $\Pi := (I - B)^{-1} \Gamma$:

$$\eta_{p+1:p+q} = \Pi \cdot \eta_{1:p} + (I - B)^{-1} \zeta, \quad (18)$$

which is a set of q regression equations. *Identifiability* is assumed, which means that B and Γ satisfy zero-constraints that allow the unambiguous recovery of the values of their free parameters from the knowledge of Π . As is well-known, this is equivalent to the specification of the ranks of certain sub-matrices of Π and to the invertibility of a certain matrix, as specified below. The observation that led to 2SLS is that:

$$\eta_{p+1:p+q} = B \cdot (\Pi \cdot \eta_{1:p}) + \Gamma \cdot \eta_{1:p} + (I - B)^{-1} \zeta \quad (19)$$

(with $\Gamma = (I - B) \cdot \Pi$). Therefore, the free elements in a row of B and Γ are regression coefficients. They can be obtained through a regression of the corresponding endogenous variable on the predicted values of the endogenous variables on the right-hand side of the equation (the relevant elements of $\Pi \cdot \eta_{1:p}$), and the exogenous variables of the equation, which are the relevant elements of $\eta_{1:p}$.

The solutions for the i^{th} row are spelled out. Let I_i select the free parameters in the i^{th} row of B (therefore, I_i is a vector containing the positions in the i^{th} row of B corresponding to free parameters), and let J_i be defined similarly for the i^{th} row of Γ . Therefore, the column vector of free parameters in the i^{th} row of B , denoted by β_i , equals

$B(i, I_i)^\top$, and for the free parameters in the i^{th} row of Γ , we define $\gamma_i := \Gamma(i, J_i)^\top$. Then:

$$\begin{aligned} \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} &= \begin{bmatrix} \text{cov}(\Pi(I_i, 1:p) \eta_{1:p}) & \text{E}(\Pi(I_i, 1:p) \eta_{1:p} \cdot \eta_{J_i}^\top) \\ & \text{cov}(\eta_{J_i}) \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} \text{E}(\Pi(I_i, 1:p) \eta_{1:p} \cdot \eta_{p+i}) \\ \text{E}(\eta_{J_i} \cdot \eta_{p+i}) \end{bmatrix}. \end{aligned} \quad (20)$$

Because of identifiability, the matrix inverted *is* invertible. Using straightforward algebra leads to the following:

$$\begin{aligned} \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} &= \begin{bmatrix} \Phi(p + I_i, 1:p) \Phi(1:p, 1:p)^{-1} \Phi(1:p, p + I_i) & \Phi(p + I_i, J_i) \\ & \Phi(J_i, p + I_i) & \Phi(J_i, J_i) \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} \Phi(p + I_i, 1:p) \Phi(1:p, 1:p)^{-1} \Phi(1:p, p + i) \\ & \Phi(J_i, p + i) \end{bmatrix}. \end{aligned} \quad (21)$$

Φ is the covariance (correlation) matrix of $\eta = [\eta_{1:p}; \eta_{p+1:p+q}]$, which is taken to be invertible (no redundant latent variables).

Equation (21) makes clear how to obtain CAN-estimators for the parameters of the structural equations: simply replace Φ by its CAN-estimator derived in the previous section. The ensuing vector with components $\hat{\beta}_i$ and $\hat{\gamma}_i$ is a smooth transformation (in the neighborhood of the true values) of $\hat{\Phi}$. Evidently, straightforward estimators (direct sample counterparts) for Π and for the covariance matrices of the residuals (for the structural form and for the reduced form) share the asymptotic properties. In fact, all parameter estimators derived so far, plus the implied estimator for Σ , are consistent and asymptotically jointly normal. They will not be asymptotically most efficient when the sample comes from a distribution like the Gaussian or an elliptical distribution: neither the weights, loadings, and correlations, nor the structural form coefficients are determined by taking all information optimally into account. The word ‘‘partial’’ has an extra connotation here. However, an advantage also exists, although this study does not elaborate on it: full-information methods are potentially vulnerable to misspecification anywhere in the system; the approach outlined here can be expected to be more robust.

Standard errors and tests-of-fit

Because of the speed of PLS, simulating the distribution of the estimators on the basis of the empirical distribution of the sample is quite feasible. Correction of bias, if any, and estimation of standard errors and confidence intervals are, in principle, well within reach. Some simulation results are reported below. Alternatively, one may use the delta-method with the Jacobian matrix calculated numerically and obtain estimates for the standard errors based on Gaussian or distribution-free asymptotic theory (either through higher-order moments or through the bootstrap).

Since the implied $\widehat{\Sigma}$ based on direct substitution of the corrected PLS estimators is consistent and asymptotically normal, if so inclined one may consider the use of overall tests as in the covariance-based SEM literature by defining a proper distance such as the trace of the square of the residual matrix $S - \widehat{\Sigma}$. When scaled by the number of observations, the distance is distributed asymptotically as a non-negative linear combination of independent $\chi^2(1)$ -variables; the coefficients are eigenvalues of a certain matrix that depends on the true parameters. One could replace them with appropriate estimates and employ a suitable approximation for the probability-value. Alternatively, and much more conveniently, the probability-value can be estimated using the bootstrap. This requires a pre-multiplication of the observation vectors by $\widehat{\Sigma}^{\frac{1}{2}}S^{-\frac{1}{2}}$, meaning that the covariance matrix of their empirical distribution satisfies the assumed (H_0) structure (see, e.g., Yuan and Hayashi, 2003, for a general discussion and elaboration in the context of covariance analysis). We will report some simulation results below.

MONTE CARLO EXPERIMENT

To assess the quality of the estimators provided by consistent PLS, a computational experiment was conducted. In particular, the performance of consistent PLS is compared with covariance-based SEM using a Monte Carlo simulation.

Setup

The first illustrative test asks for a non-trivial model that is not too large. The experiment should be challenging, meaning that the sample size should be “modest” and the number of indicators “small.” How accurate the limited information method PLS_c performs when pitted against the most efficient alternative – full information maximum likelihood (FIML) – is investigated. In some measure, the effect of non-normality will also be investigated.

The model chosen is Summers’ (1965) classical model, which is used in many econometrics studies to directly observe endogenous and exogenous variables. Here, a latent vector is composed of six components, $\eta_{1:6}$, linked to each other through the following equations:

$$\eta_{5:6} = B\eta_{5:6} + \Gamma\eta_{1:4} + \zeta, \quad (22)$$

where ζ is a two-dimensional residual vector independent of $\eta_{1:4}$, and the coefficient matrices take the following forms:

$$B := \begin{bmatrix} 0 & \beta_{12} \\ \beta_{21} & 0 \end{bmatrix} \quad (23)$$

and

$$\Gamma := \begin{bmatrix} \gamma_{11} & \gamma_{12} & 0 & 0 \\ 0 & 0 & \gamma_{23} & \gamma_{24} \end{bmatrix}. \quad (24)$$

Spelled out:

$$\eta_5 = \beta_{12}\eta_6 + \gamma_{11}\eta_1 + \gamma_{12}\eta_2 + \zeta_1 \quad (25)$$

$$\eta_6 = \beta_{21}\eta_5 + \gamma_{23}\eta_3 + \gamma_{24}\eta_4 + \zeta_2 \quad (26)$$

The endogenous variables influence one another reciprocally (for a causal interpretation of these equations see Pearl, 2009, who builds on Haavelmo, 1944). The structural form equations are not regressions but the reduced form equations are, with:

$$\Pi = \frac{1}{1 - \beta_{12}\beta_{21}} \begin{bmatrix} \gamma_{11} & \gamma_{12} & \beta_{12}\gamma_{23} & \beta_{12}\gamma_{24} \\ \beta_{21}\gamma_{11} & \beta_{21}\gamma_{12} & \gamma_{23} & \gamma_{24} \end{bmatrix}. \quad (27)$$

Of course, that $1 - \beta_{12}\beta_{21} \neq 0$ is required. All structural form coefficients can be recovered unambiguously from Π if and only if each of the submatrices $\Pi(1:2, 1:2)$ and $\Pi(1:2, 3:4)$ has rank one. Therefore, γ_{11} and γ_{12} cannot both be zero, nor can γ_{23} and γ_{24} . The following coefficients were chosen:

$$B = \begin{bmatrix} 0 & .25 \\ .50 & 0 \end{bmatrix} \quad (28)$$

and

$$\Gamma = \begin{bmatrix} -.30 & .50 & 0 & 0 \\ 0 & 0 & .50 & .25 \end{bmatrix}. \quad (29)$$

All covariances (correlations) between the exogenous latent variables $\eta_{1:4}$ are equal to .5 and the correlation between η_5 and η_6 is $\sqrt{.5}$. All numbers are chosen ‘‘arbitrarily.’’ The regression matrix equals:

$$\Pi = \begin{bmatrix} -.3429 & .5714 & .1429 & .0714 \\ -.1714 & .2857 & .5714 & .2857 \end{bmatrix}. \quad (30)$$

The correlation matrix between the latent variables can be verified as:

$$\Phi = \begin{bmatrix} 1 & .5 & .5 & .5 & .0500 & .4000 \\ .5 & 1 & .5 & .5 & .5071 & .6286 \\ .5 & .5 & 1 & .5 & .2929 & .7714 \\ .5 & .5 & .5 & 1 & .2571 & .6286 \\ .0500 & .5071 & .2929 & .2571 & 1 & \sqrt{.5} \\ .4000 & .6286 & .7714 & .6286 & \sqrt{.5} & 1 \end{bmatrix}. \quad (31)$$

Note that the correlation between η_5 and η_1 is particularly weak (.05), which may create a challenge for PLSc.

One can verify that the R-squared for the first reduced form equation is a mere .3329, whereas the corresponding value for the second equation is a healthy .7314. The implied value for the covariance matrix of ζ is:

$$\Sigma_{\zeta\zeta} = \begin{bmatrix} .5189 & -.0295 \\ -.0295 & .1054 \end{bmatrix}. \quad (32)$$

As for the λ 's, the main experiment takes just three indicators per latent variable, and all components of all loading vectors equal .70 (making their squared correlation with their corresponding latent variable less than one half).

Distributions

The model contains 31 free parameters, two from B , four from Γ , six from the correlation matrix of $\eta_{1:4}$, one correlation between ζ_1 and ζ_2 , and 18 from the loadings. All other parameters can be derived from the ones referred to (using the normalizations). Therefore, a sample size n of 300 with about 10 observations per parameter seems modest. For some experiments, $n=600$ and $n=1,200$. The leading distribution is the multivariate normal distribution. With the parameters specified as above we can determine the covariance matrix Σ of size 18×18 and generate n random drawings by $\text{randn}(n, 18) \cdot \Sigma^{\frac{1}{2}}$. The observation vectors are standardized and fed to PLS, 2SLS is employed, and outcomes noted. This process is repeated 10,000 times.

It is customary to study the effects of non-normality by the Fleishman-Vale-Maurelli procedure (see Fleishman, 1978; Vale and Maurelli, 1983). In this approach the standard normal latent variables are replaced by “well-chosen” linear combinations of powers of standard normal variables, whose correlations are such that the new latent variables have the same correlations as the original latent variables. “Well-chosen” means that specified requirements concerning the non-normal skewness and (excess-)kurtosis are satisfied. If the transformations as suggested maintain the independence between latent variables and idiosyncratic errors, then the asymptotic robustness of normal-theory statistics may apply and lead one into believing incorrectly that normality is not an issue (see Hu et al., 1992). We follow the latter authors in simply rescaling the vector of indicators by multiplying each component by the same independent random factor. Hu et al. (1992) chose $\sqrt{3} \cdot (\chi^2(5))^{-\frac{1}{2}}$, whose squared value has expectation one. This approach deliberately destroys the independence between the latent and idiosyncratic variables but

leaves Σ and the linear relationships (as well as the symmetry) undisturbed. The kurtosis of the indicators increases by six. The very same effects can be obtained by multiplying by a standard normal variable Z , which appears to yield representative samples for smaller sizes; therefore, this approach is used. In addition, multiplication is employed by a positive scale factor $\sqrt{\text{abs}(Z) \cdot \sqrt{\frac{\pi}{2}}}$ whose squared value also has expectation one. This multiplication increases the kurtosis by a lesser amount: $\frac{3\pi}{2} - 3 = 1.7124$.

Traditional PLS

Although a comparison with the traditional PLS methods is not the main goal of this paper, specifying the probability limits for the estimators of PLS mode A seems appropriate for the main parameters. Direct calculation using the true covariance matrix Σ and 2SLS for the structural parameters yields the following probability limits:

- for the loadings (regression of indicators on proxy): .8124 instead of .7;
- for the correlations between the exogenous latent variables: .3712 instead of .5;
- for the correlation between the endogenous latent variables: .5250 instead of .7071 ($= \sqrt{.5}$);
- for B : $\begin{bmatrix} 0 & .2927 \\ .5938 & 0 \end{bmatrix}$ instead of $\begin{bmatrix} 0 & .25 \\ .50 & 0 \end{bmatrix}$;
- for Γ : $\begin{bmatrix} -.1611 & .2997 & 0 & 0 \\ 0 & 0 & .3624 & .2188 \end{bmatrix}$ instead of $\begin{bmatrix} -.30 & .50 & 0 & 0 \\ 0 & 0 & .50 & .25 \end{bmatrix}$;
- for Π : $\begin{bmatrix} -.1949 & .3628 & .1284 & .0775 \\ -.1158 & .2154 & .4386 & .2648 \end{bmatrix}$ instead of $\begin{bmatrix} -.3429 & .5714 & .1429 & .0714 \\ -.1714 & .2857 & .5714 & .2857 \end{bmatrix}$.
- The implied squared correlations for the two reduced form equations are .1726 and .4421, to be compared with .3329 and .7314, respectively.

- The unrestricted regression matrix yields:

$$\begin{bmatrix} -.1705 & .3692 & .1162 & .0740 \\ -.0313 & .2386 & .4072 & .2386 \end{bmatrix}.$$

Note that the submatrices required to have rank one for identifiability in fact have rank two. In other words, the relationships that the PLS mode A proxies satisfy are at variance with the true model (and these relationships will be different again for mode B and the various modes C). This result implies that different estimation methods will yield, as a rule, different probability limits (for some of the statistical implications, see Dijkstra, 1981, 1983, 2010).

Clearly, when not corrected for inconsistency, PLS tends to give the wrong idea of the relative sizes of the parameters of the underlying covariance structure. Although the signs are correct, this is generally not guaranteed. Counterexamples for regression models are easy to provide (c.f. Dijkstra, 1981, pp. 75–76; a program generating counterexamples is available from the first author upon request).

Results

The unrestricted correlations between the latent variables

This section reports some results specific for PLSc: an estimate of the unrestricted correlation matrix of the latent variables is obtained as an intermediate product. The third and fourth columns of Table 1 contain the average and standard deviation resp. of the estimates based on 10,000 normal samples of size 300. For the fifth and sixth columns, 500 samples of size 300 were generated and 1,000 bootstrap samples were taken from each.

Evidently, for the model analyzed, PLSc yields estimators for the correlations between the latent variables that are close to unbiased. Moreover, the bootstrap can be trusted to produce relatively stable estimates of the standard errors with a slight downward bias.

Latent variable correlation	Population value	Estimates		Bootstrap std	
		mean	std	mean	std
ρ_{12}	.5000	.4993	.0639	.0639	.0048
ρ_{13}	.5000	.4990	.0639	.0640	.0054
ρ_{14}	.5000	.4997	.0640	.0641	.0054
ρ_{15}	.0500	.0535	.0798	.0792	.0051
ρ_{16}	.4000	.4020	.0691	.0682	.0050
ρ_{23}	.5000	.5000	.0654	.0639	.0051
ρ_{24}	.5000	.5006	.0649	.0640	.0053
ρ_{25}	.5071	.5060	.0645	.0633	.0052
ρ_{26}	.6286	.6286	.0581	.0567	.0053
ρ_{34}	.5000	.5004	.0650	.0642	.0052
ρ_{35}	.2929	.2951	.0724	.0715	.0050
ρ_{36}	.7714	.7709	.0480	.0474	.0052
ρ_{45}	.2571	.2590	.0740	.0726	.0049
ρ_{46}	.6286	.6293	.0581	.0572	.0054
ρ_{56}	.7071	.7049	.0530	.0520	.0054

Table 1: Unrestricted latent variable correlations obtained through PLSc.

The structural parameters

As previously stated, 2SLS is used for the “heart” of the model – the linear relationships between the latent variables. First, the differences between 2SLS and FIML are reported as applied to the latent variables *as if* they could be directly observed. Next, the analysis of the full model will reveal the size of the price to be paid for indirect observations, for both contenders.

Table 2 shows the results for 2SLS and FIML obtained from 10,000 normal samples

Parameter	True value	2SLS		FIML	
		mean	std	mean	std
γ_{11}	-.3000	-.2998	.0488	-.2996	.0487
γ_{12}	.5000	.5001	.0616	.5015	.0606
γ_{23}	.5000	.5004	.0269	.5005	.0271
γ_{24}	.2500	.2500	.0244	.2498	.0242
β_{12}	.2500	.2501	.0755	.2480	.0752
β_{21}	.5000	.5001	.0423	.5001	.0431
convergence:		100.00 %		99.64 %	

Table 2: 2SLS and FIML applied to the true latent variable scores.

of size 300 (9,964 samples for FIML attributable to 36 cases of non-convergence). As is shown, the FIML and the 2SLS estimators are virtually unbiased. The results are very similar, and for this particular model there does not appear to be much to choose from between the two competitors.

What happens if the variables are no longer observable? Specifically, what happens when we have only three modestly correlated indicators per latent variable?

We look at the leading case – multivariate normality – first. Table 3 presents the FIML results for 9982 samples (for 18 samples, FIML did not converge). Some slight bias in the estimates is visible, and the price for unobservability is quite substantial: roughly, the standard errors are doubled or tripled. Table 3 also presents the PLS_c results. Again, as in the case of the observable variables, a very similar performance of PLS_c and FIML is observed. We note that traditional PLS using 2SLS for the structural model produces relatively stable estimators, approximately unbiased for *its* probability limits (which deviate strongly from the true values). Recall that given the misrepresentation of the correlation structure of the latent variables by traditional PLS, other simultaneous

Parameter	True value	FIML		PLSc		PLS (with 2SLS)	
		mean	std	mean	std	mean	std
γ_{11}	−.3000	−.3045	.0920	−.2990	.0905	−.1615	.0549
γ_{12}	.5000	.5103	.1263	.4994	.1155	.2994	.0681
γ_{23}	.5000	.5027	.0818	.5002	.0751	.3619	.0509
γ_{24}	.2500	.2496	.0718	.2502	.0732	.2188	.0493
β_{12}	.2500	.2452	.1367	.2526	.1315	.2944	.1073
β_{21}	.5000	.5040	.1390	.4983	.1323	.5927	.1335
convergence:		99.82 %		100.00 %		100.00 %	

Table 3: Structural model results for 300 observations, multivariate normality (estimates based on 10,000 simulation samples; standard errors based on 500 simulation samples with 1,000 bootstrap samples each).

equation estimation methods (and other modes) will yield different central values for the estimators. In the sequel, we focus on PLSc and FIML.

Next, the performance of FIML and PLSc in the case of moderately non-normal data is examined. If the indicators are scaled by the factor $\sqrt{\text{abs}(Z) \cdot \sqrt{\frac{\pi}{2}}}$, which means that the kurtosis is increased by a modest 1.7124, both methods suffer but again in similar ways. Table 4 shows the results for FIML and PLSc. Again, both estimators are of similar quality.

Finally, the performance of FIML and PLSc in the case of highly kurtotic data is examined. Specifically, the indicators are rescaled by the normal variable Z , increasing the kurtosis to six. Now, for the first time, PLSc showed some apparent convergence problems for two out of 10,000 samples; in 96.7% of the cases, PLSc converged in no more than seven steps. FIML struggled a bit more: for 1,067 out of 10,000 samples, convergence did not occur. Table 5 presents the results for FIML and PLSc. Again,

Parameter	True value	FIML		PLSc	
		mean	std	mean	std
γ_{11}	-.3000	-.3023	.1134	-.2976	.1133
γ_{12}	.5000	.5097	.1502	.4987	.1472
γ_{23}	.5000	.5018	.0924	.4998	.0949
γ_{24}	.2500	.2492	.0879	.2478	.0926
β_{12}	.2500	.2412	.1703	.2550	.1675
β_{21}	.5000	.5040	.1780	.5015	.1686
convergence:		98.42 %		100.00 %	

Table 4: Structural model results for 300 observations, kurtosis=1.7124 (estimates based on 10,000 simulation samples; standard errors based on 500 simulation samples with 1,000 bootstrap samples each).

Parameter	True value	FIML		PLSc	
		mean	std	mean	std
γ_{11}	-.3000	-.3108	.1612	-.2976	.1604
γ_{12}	.5000	.5289	.2118	.5018	.2101
γ_{23}	.5000	.5063	.1319	.4970	.1338
γ_{24}	.2500	.2516	.1245	.2501	.1317
β_{12}	.2500	.2306	.2449	.2574	.2362
β_{21}	.5000	.4949	.2489	.4991	.2422
convergence:		89.33 %		99.98 %	

Table 5: Structural model results for 300 observations, kurtosis=6 (estimates based on 10,000 simulation samples; standard errors based on 500 simulation samples with 1,000 bootstrap samples each).

both techniques exhibit a comparable performance. The FIML estimates appear slightly more biased than PLSc.

Rescaling the indicator vectors by $\sqrt{\text{abs}(Z) \cdot \sqrt{\frac{\pi}{2}}}$ leads to standard errors for *both* methods that are approximately $\sqrt{3\pi/6} \approx 1.25$ times as large as for the normal case. Rescaling by Z effectively multiplies the normal case standard errors by approximately $\sqrt{3} \approx 1.73$. These numbers are no accident; they agree with the (asymptotic) corrections for non-normal kurtosis in covariance structure analysis (c.f. Bentler and Dijkstra, 1985). Therefore, the ratios between corresponding standard errors of PLSc and FIML are constant for the analyzed conditions.

It seems fair to say for this particular setup that both “contenders” tend to produce estimators for the structural parameters of “the same or comparable quality”. Of course, under non-normality (and especially with non-independence between errors and latent variables), one cannot be sure that the usual way to calculate standard errors (as derived from the information matrix) yields proper estimates. With FIML, one could anticipate underestimation for the non-normal models with the increased kurtosis considered here, as borne out by some experiments (not displayed). One should use one of the known correction methods available in standard software such as EQS or, simply the bootstrap. The latter is the obvious choice for PLSc.

The loadings

Based on 10,000 *normal* samples (9982 for FIML), the 18 loading estimators are essentially unbiased for both methods. The FIML average values differ from .7000 by no more than .0010 (and by .0005 on average), and for PLSc these numbers are .0040 and .0020, respectively. The FIML standard errors are definitely smaller. Because they appear to be equal for the loadings on the same latent variable, and the same is true for PLSc, averages of standard errors per latent variable are reported (see Table 6).

These results are reinforced for the non-normal distributions. Unbiasedness is not

Latent Variable	std of loadings	
	FIML	PLSc
η_1	.0426	.0670
η_2	.0415	.0507
η_3	.0405	.0524
η_4	.0420	.0569
η_5	.0409	.0734
η_6	.0365	.0403

Table 6: Average standard errors of loadings per latent variable.

affected but the standard errors are. This paper does not report them because the same phenomenon occurs as for the structural form parameters (and even more clearly). Rescaling the indicator vectors by $\sqrt{\text{abs}(Z)} \cdot \sqrt{\frac{\pi}{2}}$ leads again to standard errors for *both* methods that are all approximately $\sqrt{3\pi/6} \approx 1.25$ times as large as for the normal case. Additionally, rescaling by Z effectively multiplies the normal case standard errors by $\sqrt{3} \approx 1.73$. Therefore, the ratios between the corresponding standard errors of PLSc and FIML are also constant for the analyzed conditions.

With three modestly correlated indicators per latent variable, the full information method lives up to its expectations as far as the loadings are concerned. This situation may change with a more unfavorable ratio of observations to indicators (doubling the indicators, for example) but has yet to be pursued. This study has also not attempted to experiment with the simple PLSc algorithm using other selections of latent variables in the regressions leading to the sample proxies, or changing the coefficients in those regressions. Clearly, additional work is required.

Overall test-of-fit

The experiment also provided insights into the performance of the overall test-of-fit. Here some bootstrap results are reported for the normal distribution only. In the case of non-normal kurtosis values, one may need to “downweight” the observations, as in Yuan and Hayashi (2003), which may require delicate fine-tuning. Because this first study is meant to be illustrative of the possibilities of PLSc, a much more elaborate analysis is postponed to another occasion. In addition we just investigate the relative frequency with which the true model is rejected when the fit statistic exceeds the bootstrap-based estimate of a conventional quantile, as well as the distribution of the probability values (the relative frequency with which the bootstrapped distances exceed the observed distance). Power analyses are deferred to the future. Yuan and Bentler (1998) and Yuan and Hayashi (2003) should be reviewed for careful discussions of the issues involved in testing whether a covariance model fits.

Two distance functions are considered:

$$d_{LS} := \frac{1}{2} \text{trace} \left(S - \widehat{\Sigma} \right)^2, \quad (33)$$

the squared Euclidean distance, and

$$d_G := \frac{1}{2} \sum_{k=1}^{\#\text{indicators}} (\log(\varphi_k))^2, \quad (34)$$

the “geodesic” distance. Here, φ_k is the k^{th} eigenvalue of $S^{-1}\widehat{\Sigma}$. Both distances are zero if and only if the model fits perfectly: $\widehat{\Sigma} = S$. They belong to different classes: d_{LS} cannot be expressed in terms of the eigenvalues. The geodesic distance is one of Swain’s (1975) fitting functions; normalized, they are asymptotically equivalent to the likelihood ratio statistic. d_G is characterized by the property that its minimization with respect to the parameters leaves the generalized variance intact (given scale invariance of the model, see Dijkstra, 1990). When evaluated at PLSc’s $\widehat{\Sigma}$, of course it cannot be expected to follow a χ^2 -distribution.

Observations	Rejection probabilities		
	Nominal	d_G	d_{LS}
300	10.0 %	4.2 %	4.9 %
	5.0 %	1.5 %	1.8 %
	2.5 %	0.5 %	0.5 %
600	10.0 %	7.7 %	8.2 %
	5.0 %	3.9 %	4.0 %
	2.5 %	1.8 %	1.7 %
1200	10.0 %	9.1 %	10.3 %
	5.0 %	5.0 %	4.3 %
	2.5 %	2.8 %	1.9 %

Table 7: Rejection probabilities of two discrepancy functions for PLSc.

We generated 1,000 normal samples of size 300. For each sample, the implied correlation matrix $\widehat{\Sigma}$ and its distance to the sample correlation matrix S , $d(\widehat{\Sigma}, S)$ were calculated, both for least squares and the geodesic distance. The observation vectors were pre-multiplied by $\widehat{\Sigma}^{\frac{1}{2}}S^{-\frac{1}{2}}$ and 1,000 bootstrap samples of the transformed values were generated, with the model refitted and distances re-calculated for each bootstrap sample. We note whether the observed distance $d(\widehat{\Sigma}, S)$ exceeds certain quantiles of the empirical distribution function of the bootstrapped distances. A false alarm occurs if it does. Ideally, the average number of false alarms agrees with the theoretical values.

The upper third of Table 7 contains the empirical rejection probabilities for several levels of nominal rejection probabilities based on 300 observations. These are not too bad as things go (see Yuan and Bentler, 1998; Yuan and Hayashi, 2003), but they are clearly too small. The test is more cautious than desired. Apparently, when the sample does not fit too well, the bootstrapped transformed sample tends to be worse. Also (not

shown), the histogram of the probability values is not uniform, and its shape is like a parabola with the maximum value in the middle. Increasing the sample size helps. The same exercise for $n = 600$ results in the values reported in the middle part of Table 7. Now the histogram of the probability values is definitely closer to uniform. The lower part of Table 7 shows that by doubling again, $n = 1200$, the rejection probabilities are almost right. Again, the histogram of probability values is closer to uniform. The balanced bootstrap yielded very similar results.

Summary of experimental results

For the specific model analyzed in this section, PLSc provides as good a picture of the all-important structural parameters as FIML does. In addition, the unrestricted correlations allow of proper estimation, and there are grounds to believe a correct test-of-fit is well within reach. The loadings are more difficult to estimate accurately; here, FIML has a definite advantage. However, on the whole, one can certainly maintain that the results are rather encouraging.

LIMITATIONS AND AVENUES FOR FUTURE RESEARCH

The point of departure in this paper is a “clean” second-order factor model (‘LISREL’) in which indicators are, in principle, mutually correlated and correlated with all latent variables, but each indicator loads directly only on one latent variable. This is what Wold called the “basic design.” In practice, one may encounter situations that are conceptually less clear-cut, and loading matrices as well as covariance matrices of errors of different latent variables may have structures more complicated than handled in this paper. How these situations affect PLSc is as yet unclear, but it certainly requires analysis. Other topics for further research were already alluded to in the text. These topics include power analyses for overall tests-of-fit and tests of robustness against structural misspecification to be compared with full information methods. Obviously, the power analyses and

robustness tests are related, and they are best studied in conjunction, allowing a proper assessment of the expected trade-off between the two. An investigation into the effects of skewness and kurtosis on convergence, speed, and stability of the algorithm will also be interesting.

Existing simulation studies on the performance of PLS with respect to reflective measurement should be replicated using consistent PLS. Many characteristics that were already explored for PLS, such as the statistical power (Chin and Newsted, 1999; Goodhue et al., 2006), parameter accuracy and convergence behavior (Reinartz et al., 2009; Henseler, 2010), or performance relative to other component-based structural equation modeling techniques (Hwang et al., 2010; Lu et al., 2011; McDonald, 1996; Tenenhaus, 2008), need to be investigated for consistent PLS.

For PLSc, 2SLS was suggested for the relationships between the latent variables, but many alternatives exist. A class of alternatives that suggests itself is the set of minimum distance or GLS estimators based on unrestricted CAN-estimators for the regression matrix. They will probably be computationally more expensive and less robust, but perhaps more efficient (asymptotically) under correct specification.

Moreover, note that because PLSc yields CAN-estimators, *one* Gauss-Newton iteration on the first-order conditions for ML, starting from PLSc, will yield asymptotically efficient estimators equivalent to ML, provided that the model is correctly specified (see Bentler and Dijkstra, 1985, for the general theory). For this new estimator, the same research topics suggest themselves. This new estimator is expected to be less robust against misspecifications than PLSc as developed in this paper.

Finally, future research should extend consistent PLS to nonlinear latent variable models with interaction terms, squares, and higher order powers. Here, consistency with numerical expediency is difficult to obtain with covariance-based structural equation modeling (see Schumacker and Marcoulides, 1998; Klein and Muthén, 2007, for an overview and discussion), whereas PLS has shown to be a viable alternative (Chin et al.,

2003; Henseler and Chin, 2010; Goodhue et al., 2007). PLS is a rather promising platform for this endeavor (c.f. Dijkstra and Schermelleh-Engel, 2012) and deserves further investigation.

References

- Bekker, P. A., Dijkstra, T. K., 1990. On the nature and number of the constraints on the reduced form as implied by the structural form. *Econometrica* 58 (2), 507–514.
- Bekker, P. A., Merckens, A., Wansbeek, T. J., 1994. *Identification, Equivalent Models, and Computer Algebra*. Academic Press, Boston.
- Bentler, P. M., Dijkstra, T. K., 1985. Efficient estimation via linearization in structural models. In: Krishnaiah, P. R. (Ed.), *Multivariate Analysis VI*. North-Holland, Amsterdam, pp. 9–42.
- Boardman, A., Hui, B., 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.
- Chin, W. W., 1998. Issues and opinion on structural equation modeling. *MIS Quarterly* 22 (1), vii–xvi.
- Chin, W. W., Dibbern, J., 2010. An introduction to a permutation based procedure for multi-group PLS analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In: Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. Vol. II of *Computational Statistics*. Springer, Heidelberg, Dordrecht, London, New York, pp. 171–193.
- Chin, W. W., Marcolin, B. L., Newsted, P. R., 2003. A partial least squares latent variable modeling approach for measuring interaction effects. Results from a Monte

- Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research* 14 (2), 189–217.
- Chin, W. W., Newsted, P. R., 1999. Structural equation modeling analysis with small samples using partial least squares. In: Hoyle, R. H. (Ed.), *Statistical Strategies for Small Sample Research*. Sage Publications, Thousand Oaks, London, New Delhi, Ch. 12, pp. 307–341.
- Chin, W. W., Thatcher, J. B., Wright, R. T., forthcoming. Assessing common method bias: Problems with the ULMC technique. *MIS Quarterly*.
- Dijkstra, T. K., 1981. Latent variables in linear stochastic models: Reflections on "Maximum Likelihood" and "Partial Least Squares" methods. Ph.D. thesis, Groningen University, Groningen, a second edition was published in 1985 by Sociometric Research Foundation.
- Dijkstra, T. K., 1983. Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics* 22 (1/2), 67–90.
- Dijkstra, T. K., 1990. Some properties of estimated scale invariant covariance structures. *Psychometrika* 55 (2), 327–336.
- Dijkstra, T. K., 2010. Latent variables and indices: Herman Wold's basic design and partial least squares. In: Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. Vol. II of *Computational Statistics*. Springer, Heidelberg, Dordrecht, London, New York, pp. 23–46.
- Dijkstra, T. K., Schermelleh-Engel, K., 2012. Consistent partial least squares for non-linear structural equation models. *Submitted*.
- Fleishman, A., 1978. A method for simulating non-normal distributions. *Psychometrika* 43 (4), 521–532.

- Fornell, C., Bookstein, F. L., 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research* 19 (4), 440–452.
- Gefen, D., Rigdon, E. E., Straub, D., 2011. An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly* 35 (2), iii–xiv.
- Goodhue, D., Lewis, W., Thompson, R., 2006. PLS, small sample size, and statistical power in MIS research. In: HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on System Sciences. IEEE, p. 202b.
- Goodhue, D., Lewis, W., Thompson, R., 2007. Statistical power in analyzing interaction effects: Questioning the advantage of PLS with product indicators. *Information Systems Research* 18 (2), 211–227.
- Haavelmo, T., 1944. The probability approach in econometrics. *Econometrica* 12 (Supplement), iii–vi, 1–115.
- Hair, J. F., Sarstedt, M., Ringle, C. M., Mena, J. A., 2012. An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science* 40 (3), 414–433.
- Henseler, J., 2010. On the convergence of the partial least squares path modeling algorithm. *Computational Statistics* 25 (1), 107–120.
- Henseler, J., 2012. Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science* 40 (3), 402–413.
- Henseler, J., Chin, W. W., 2010. A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling* 17 (1), 82–109.

- Henseler, J., Fassott, G., Dijkstra, T. K., Wilson, B., 2012. Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. *European Journal of Information Systems* 21 (1), 99–112.
- Henseler, J., Sarstedt, M., 2012. Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, DOI: 10.1007/s00180-012-0317-1.
- Hsieh, J. J., Rai, A., Keil, M., 2008. Understanding digital inequality: comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. *MIS Quarterly* 32 (1), 97–126.
- Hu, L.-t., Bentler, P. M., Kano, Y., 1992. Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin* 112 (2), 351–362.
- Hwang, H., Malhotra, N., Kim, Y., Tomiuk, M., Hong, S., 2010. A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research* 47 (4), 699–712.
- Keil, M., Tan, B., Wei, K., Saarinen, T., Tuunainen, V., Wassenaar, A., 2000. A cross-cultural study on escalation of commitment behavior in software projects. *MIS Quarterly* 24 (2), 299–325.
- Klein, A., Muthén, B., 2007. Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research* 42 (4), 647–673.
- Liang, H., Saraf, N., Hu, Q., Xue, Y., 2007. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *MIS Quarterly* 31 (1), 59–87.
- Lohmöller, J. B., 1989. *Latent Variable Path Modeling with Partial Least Squares*. Physica, Heidelberg.

- Lu, I., Kwan, E., Thomas, D., Cedzynski, M., 2011. Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing* 28 (3), 258–268.
- Marcoulides, G. A., Saunders, C., 2006. PLS: A silver bullet? *MIS Quarterly* 30 (2), iii–ix.
- McDonald, R., 1996. Path analysis with composite variables. *Multivariate Behavioral Research* 31 (2), 239–270.
- Pearl, J., 2009. *Causality: models, reasoning, and inference*, 2nd Edition. Cambridge University Press.
- Qureshi, I., Compeau, D., 2009. Assessing between-group differences in information systems research: A comparison of covariance- and component-based SEM. *MIS Quarterly* 33 (1), 197–214.
- Reinartz, W., Haenlein, M., Henseler, J., 2009. An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing* 26 (4), 332–344.
- Ringle, C. M., Sarstedt, M., Straub, D. W., 2012. A critical look at the use of PLS-SEM in *MIS Quarterly*. *MIS Quarterly* 36 (1), iii – xiv.
- Schumacker, R. E., Marcoulides, G. A. (Eds.), 1998. *Interaction and Nonlinear Effects in Structural Equation Modeling*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Summers, R., 1965. A capital intensive approach to the small sample properties of various simultaneous equation estimators. *Econometrica* 33 (1), 1–41.
- Swain, A., 1975. A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika* 40 (3), 315–335.

- Tenenhaus, M., 2008. Component-based structural equation modelling. *Total Quality Management & Business Excellence* 19 (7-8), 871–886.
- Tenenhaus, M., Vinzi, V. E., Chatelin, Y. M., Lauro, C., 2005. PLS path modeling. *Computational Statistics & Data Analysis* 48 (1), 159–205.
- Vale, C., Maurelli, V., 1983. Simulating multivariate nonnormal distributions. *Psychometrika* 48 (3), 465–471.
- Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H. (Eds.), 2010. *Handbook of Partial Least Squares: Concepts, Methods, and Applications*. Vol. II of *Computational Statistics*. Springer, Heidelberg, Dordrecht, London, New York.
- Wetzels, M., Odekerken-Schröder, G., van Oppen, C., 2009. Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly* 33 (1), 177–195.
- Wold, H. O. A., 1966. Non-linear estimation by iterative least squares procedures. In: David, F. N. (Ed.), *Research Papers in Statistics*. Wiley, London, New York, Sydney, pp. 411–444.
- Wold, H. O. A., 1975. Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. In: Gani, J. (Ed.), *Perspectives in Probability and Statistics*. Papers in honour of M. S. Bartlett on the occasion of his sixty-fifth birthday. Applied Probability Trust, Sheffield, pp. 117–142.
- Wold, H. O. A., 1982. Soft modelling: the basic design and some extensions. In: Jöreskog, K. G., Wold, H. O. A. (Eds.), *Systems under Indirect Observation. Causality, Structure, Prediction*. Vol. II. North-Holland, Amsterdam, New York, Oxford, pp. 1–54.
- Yuan, K., Bentler, P., 1998. Normal theory based test statistics in structural equation modelling. *British Journal of Mathematical and Statistical Psychology* 51 (2), 289–309.

Yuan, K., Hayashi, K., 2003. Bootstrap approach to inference and power analysis based on three test statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology* 56 (1), 93–110.