

Composites as Factors: canonical variables revisited
working paper, Theo K. Dijkstra, June-22-2013

In this note we specify a factor model with observables only. Its parameters are estimated in a natural way by generalized canonical variables programs. The paper is meant to contribute to the ongoing discussion about the merits of Partial Least Squares.

0. Introduction

Canonical variables attempt to capture in low dimensions the information that separate sets of measured variables contain about each other. There need not be an underlying model, no specification of a hypothetical generator, partly random, partly structural, of the observations. The researcher could simply be involved in an exploratory analysis of a high-dimensional dataset, using a partition of the observed variables, that is neither definitive nor completely arbitrary. Canonical variables can be of great help there.

But there are also situations where one is willing to entertain a path diagram, that maps possible dependencies between concepts as measured or defined in a *formative* way. Observed variables, ‘indicators’, determine the content of the concepts. One of the goals of the analysis is to assign weights to the indicators for the concepts, to construct ‘indices’ or ‘linear composites’ that honor their interrelationships as well as possible. So we have a ‘measurement triplet’: a conceptual structure, as embodied in the path diagram, the specified indicators, and their composites. We will adopt a modeling principle from PLS¹: all information between the blocks is conveyed by the composites. The next section specifies the implied model, it is similar to yet clearly distinct from a classic factor model. In section two we will show that classical generalized canonical variable estimators are capable of retrieving the underlying parameters from a knowledge of the covariance/correlation matrix. This is true also for PLS’ mode B, as well as mode A (with a simple modification). When applied to a sample covariance/correlation the resulting estimators are consistent and asymptotically normal, under standard conditions (a Gaussian distribution is allowed of course but not required). The third section outlines possible ways of testing the composite factor model, using overall goodness-of-fit criteria and ‘local tests’. Section four concludes and suggests further extensions of the model.

¹In PLS this is seen as a basic principle in soft modeling. See e.g. Wold’s chapter on ‘Soft Modeling: The Basic Design and Some Extensions’ in K.G. Jöreskog & H. Wold (1982), (eds.), *Systems under indirect observation, part II*, North-Holland, Amsterdam.

1. The composites factor model

Suppose we have a zero mean random vector x , decomposable into N sub-vectors x_1, x_2, \dots, x_N , with Σ , the covariance/correlation matrix² of x decomposed conformably:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1N} \\ & \Sigma_{22} & \dots & \Sigma_{2N} \\ & & \dots & \dots \\ & & & \Sigma_{NN} \end{bmatrix}. \quad (1)$$

Here each $\Sigma_{ii} := E x_i x_i^\top$ is assumed to be positive definite, and for all $i \neq j$ we have

$$\Sigma_{ij} := E x_i x_j^\top = \rho_{ij} \Sigma_{ii} w_i \cdot w_j^\top \Sigma_{jj} \quad (2)$$

where each weight vector w_i is normalized: $\text{var}(w_i^\top x_i) = w_i^\top \Sigma_{ii} w_i = 1$. We will take $\Psi := (\rho_{ij})$ positive definite. This, together with all $\Sigma_{ii} > 0$ entails that Σ itself is positive definite. Note that because of the normalization $w_i^\top \Sigma_{ij} w_j = \text{corr}(w_i^\top x_i, w_j^\top x_j) = \rho_{ij}$. In addition, $\text{corr}(x_i, w_i^\top x_i) = \Sigma_{ii} w_i$. So according to the model the ‘cross-covariances’ Σ_{ij} are completely determined by the correlations between the linear composites and their correlations with their direct indicators. In fact we have

$$\Sigma_{ij} = \text{corr}(w_i^\top x_i, w_j^\top x_j) \cdot \text{corr}(x_i, w_i^\top x_i) \cdot \text{corr}(x_j^\top, w_j^\top x_j) \quad (3)$$

There are serious restrictions on the ranks of the off-diagonal matrices. In particular, every sub-matrix of the form $[\Sigma_{i,i+1}, \dots, \Sigma_{iN}]$ has rank one, and similarly for ‘columns’.

If we define for every block a vector of ‘measurement errors’

$$\delta_i := x_i - \Sigma_{ii} w_i \cdot w_i^\top x_i \quad (4)$$

(perhaps a better name would be ‘redundancy vector’ since it separates the information from x_i that is uncorrelated with its composite), it is easily verified that they possess the following properties (with $i \neq j$):

$$E(\delta_i \cdot w_i^\top x_i) = 0 \quad (5)$$

$$E(\delta_i \cdot w_j^\top x_j) = 0 \quad (6)$$

$$E(\delta_i \delta_j^\top) = 0. \quad (7)$$

²Generally we will assume that all indicators are standardized, with a zero mean and unit standard deviation.

In words: the measurement vectors are uncorrelated with every composite, and uncorrelated between blocks. This is rather reminiscent of a classical factor model, in particular the ‘basic design’ in PLS, but it is certainly not identical with it. The factor model adds N unobserved latent variables to x , one for each x_i . These latent variables replace the composites. The ensuing covariance matrix has the same rank-restrictions as Σ on the off-diagonal sub-matrices. But usually the matrices on the diagonal satisfy additional restrictions, due to zero-correlation assumptions concerning the measurement errors *within* the blocks. At least some zero-correlations are needed to identify the loadings. A typical, strong assumption is that the covariance matrices of the measurement vectors are diagonal. This is not possible for δ_i since clearly $w_i^\top \delta_i$ is identically zero: so the elements of δ_i can only be uncorrelated when they are all zero with probability one, and Σ_{ii} , in that case with rank one, would not be invertible. It may sometimes be possible to rescale the vectors, by multiplying $\Sigma_{ii} w_i$ by a scalar c_i , such that $(\rho_{ij}/(c_i c_j))$ and the ‘error covariance matrices’ $\Sigma_{ii} - c_i^2 \cdot \Sigma_{ii} w_i w_i^\top \Sigma_{ii}$ are positive definite with at least some off-diagonal zeros. Then one could re-interpret Σ as the covariance matrix for a traditional factor model, at the expense of introducing N additional unobserved variables. But this does not seem necessary³, and Occam *cum suis* would advice against it.

So far we have not restricted the values of the ρ_{ij} , they could be anything provided $(\rho_{ij}) > 0$. In simulation studies one may wish to work with fixed values for the regression coefficients and the multiple correlation coefficient. For three blocks e.g. one could want particular values for β_1 , β_2 and $\rho_{3.12}^2$ respectively. Since

$$\beta_1 + \rho_{12}\beta_2 = \rho_{13} \quad (8)$$

$$\rho_{12}\beta_1 + \beta_2 = \rho_{23} \quad (9)$$

and

$$\beta_1^2 + \beta_2^2 + 2\beta_1\beta_2\rho_{12} = \rho_{3.12}^2 \quad (10)$$

we can solve the last equation for ρ_{12} , and then use the previous two for ρ_{13} and ρ_{23} . It remains to check for positive definiteness that $\rho_{12}^2 < 1$ and

$$\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} < 1. \quad (11)$$

The weight vectors are perfectly free (apart from the normalization of course) as well as the (p.d.) covariance matrices Σ_{ii} . So if we generate a random

³The additional factor model structure on Σ is not necessarily a bad thing. Predictions based on estimators incorporating the constraints, with the implied reduction of the number of parameters, even when or perhaps just because they are incorrect, might well be more accurate than those not using them.

vector x with Σ as specified, we know that the composites obey the required regression equation

$$w_3^\top x_3 = \beta_1 \cdot w_1^\top x_1 + \beta_2 \cdot w_2^\top x_2 + \epsilon_{3.12} \quad (12)$$

where the implied residual $\epsilon_{3.12}$ is by construction uncorrelated with the explanatory composites, and the multiple R^2 equals $\rho_{3.12}^2$. Note that the best (least squares) predictor \hat{x}_3 of x_3 in terms of x_1 and x_2 is simply:

$$\hat{x}_3 = \Sigma_{33} w_3 \cdot (\beta_1 \cdot w_1^\top x_1 + \beta_2 \cdot w_2^\top x_2) \quad (13)$$

$$= \left[\Sigma_{33} w_3 \cdot \begin{bmatrix} \beta_1 \cdot w_1^\top & \beta_2 \cdot w_2^\top \end{bmatrix} \right] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \quad (14)$$

A regression prediction of an indicator of the third block based on x_1 and x_2 is the product of its correlation with the third composite and the prediction of the latter in terms of the other composites. The regression matrix has rank one.

Models with rank-restrictions on sub-matrices of (ρ_{ij}) , as is the case for interdependent simultaneous equation systems, will require some more work but no more than for traditional factor models.

2. Generalized canonical variables

Optimization programs designed for the extraction of canonical variables appear to have a natural role here. The special case of two blocks, treated extensively in standard multivariate literature⁴, is the least interesting, since all alternative programs are then identical. Here we will discuss a number of the generalizations as developed for three and more blocks⁵, and show that they are all capable of retrieving the parameters from Σ . The implication, provable with some standard asymptotic arguments, is that when applied to S they will all produce CAN-estimators when S is CAN for Σ . If the model is incorrect, so that $\text{plim}(S) \neq \Sigma$, they will tend to different probability limits, and their sample differences can be used for testing (section 3). For notational ease we will use a composites factor model with three blocks, $N = 3$, but that is no limitation.

Now consider the covariance matrix, denoted by $\Psi(v)$, of $v_1^\top x_1$, $v_2^\top x_2$ and

⁴See e.g. chapter twelve from T. W. Anderson (1984), *Introduction to Multivariate Statistical Analysis*, Wiley, New York.

⁵The reference here is J. R. Kettenring (1971). Canonical analysis of several sets of variables, *Biometrika*, 58, 3, 433-451. See also chapter three from R. Gnanadesikan (1997). *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York.

$v_3^\top x_3$ where each v_i is normalized ($\text{var}(v_i^\top x_i) = 1$). So

$$\Psi(v) := \begin{bmatrix} 1 & v_1^\top \Sigma_{12} v_2 & v_1^\top \Sigma_{13} v_3 \\ v_1^\top \Sigma_{12} v_2 & 1 & v_2^\top \Sigma_{23} v_3 \\ v_1^\top \Sigma_{13} v_3 & v_2^\top \Sigma_{23} v_3 & 1 \end{bmatrix}. \quad (15)$$

Canonical variables are composites whose correlation matrix has ‘maximum distance’ to the identity matrix of the same size. They are ‘collectively maximally correlated’. The term is clearly ambiguous for more than two blocks. One program that would seem to be natural is to maximize with respect to v

$$z(v) := \text{abs}(\Psi_{12}) + \text{abs}(\Psi_{13}) + \text{abs}(\Psi_{23}) \quad (16)$$

subject to the usual normalizations. Since

$$\text{abs}(\Psi_{ij}) = \text{abs}(\rho_{ij}) \cdot \text{abs}(v_i^\top \Sigma_{ii} w_i) \cdot \text{abs}(v_j^\top \Sigma_{jj} w_j) \quad (17)$$

we know, thanks to Cauchy-Schwarz, that

$$\text{abs}(v_i^\top \Sigma_{ii} w_i) = \text{abs}\left(v_i^\top \Sigma_{ii}^{\frac{1}{2}} \Sigma_{ii}^{\frac{1}{2}} w_i\right) \leq \sqrt{v_i^\top \Sigma_{ii}^{\frac{1}{2}} \Sigma_{ii}^{\frac{1}{2}} v_i \cdot w_i^\top \Sigma_{ii}^{\frac{1}{2}} \Sigma_{ii}^{\frac{1}{2}} w_i} \quad (18)$$

$$= \sqrt{v_i^\top \Sigma_{ii} v_i \cdot w_i^\top \Sigma_{ii} w_i} = 1 \quad (19)$$

with equality if and only if $v_i = w_i$ (ignoring irrelevant sign differences). Observe that the upper bound can be reached for $v_i = w_i$ for all terms in which v_i appears, so maximization of the sum of the absolute correlations gives w . A numerical, iterative routine⁶ suggests itself by noting that the optimal v_1 satisfies the first order condition

$$0 = \text{sgn}(\Psi_{12}) \cdot \Sigma_{12} v_2 + \text{sgn}(\Psi_{13}) \cdot \Sigma_{13} v_3 - l_1 \Sigma_{11} v_1 \quad (20)$$

where l_1 is a Lagrange multiplier (for the normalization), and two other quite similar equations for v_2 and v_3 . So with arbitrary starting vectors one could solve the equations recursively for v_1 , v_2 and v_3 respectively, updating them after each full round or at the first opportunity, until they settle down at the optimal value. Note that each update of v_1 is obtainable by a regression of a ‘sign-weighted sum’

$$\text{sgn}(\Psi_{12}) \cdot v_2^\top x_2 + \text{sgn}(\Psi_{13}) \cdot v_3^\top x_3 \quad (21)$$

⁶With Σ one does not really need an iterative routine of course: $\Sigma_{ij} = \rho_{ij} \Sigma_{ii} w_i w_j^\top \Sigma_{jj}$ can be solved directly for the weights (and the correlation). But in case we just have S , an algorithm comes in handy.

on x_1 , and analogously for the other weights. This happens to be the classical form of PLS' mode B⁷. For Σ we do not need many iterations, to put it mildly: the update of v_1 is already w_1 , as straightforward algebra will easily show. And similarly for the other weight vectors. In other words, we have in essentially just one iteration a fixed point for the mode B equations that is precisely w .

If we use the correlations themselves in the recursions instead of just their signs, we regress the ‘correlation weighted sum’

$$\Psi_{12} \cdot v_2^\top x_2 + \Psi_{13} \cdot v_3^\top x_3 \quad (22)$$

on x_1 (and analogously for the other weights), and end up with weights that maximize

$$z(v) := \Psi_{12}^2 + \Psi_{13}^2 + \Psi_{23}^2, \quad (23)$$

the simple sum of the squared correlations. Again, with the same argument, the optimal value is w .

Observe that for this $z(v)$ we have

$$\text{tr}(\Psi^2) = 2 \cdot z(v) + 3 = \sum_{i=1}^3 \gamma_i^2 \quad (24)$$

where $\gamma_i := \gamma_i(\Psi(v))$ is the i^{th} eigenvalue of $\Psi(v)$. We can take other functions of the eigenvalues, in order to maximize the difference between $\Psi(v)$ and the identity matrix of the same order. Kettenring (1971) discusses a number of alternatives. One of them minimizes $\prod_i \gamma_i$, the determinant of $\Psi(v)$, also known as the generalized variance, first suggested by Steel (1951)⁸. The program is called ‘genvar’. Since $\sum_i \gamma_i$ is always N (three in this case) for every choice of v , genvar tends to make the eigenvalues as diverse as possible (as opposed to the identity matrix where they are all equal to one). Note that the determinant of Ψ equals $(1 - \Psi_{23}^2)$ times

$$\begin{aligned} & 1 - \begin{bmatrix} \Psi_{12} & \Psi_{13} \end{bmatrix} \begin{bmatrix} 1 & \Psi_{23} \\ \Psi_{23} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \Psi_{12} \\ \Psi_{13} \end{bmatrix} \\ &= 1 - (v_1^\top \Sigma_{11} w_1)^2 \cdot \begin{bmatrix} \rho_{12} v_2^\top \Sigma_{22} w_2 & \rho_{13} v_3^\top \Sigma_{33} w_3 \end{bmatrix} \begin{bmatrix} 1 & \Psi_{23} \\ \Psi_{23} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{12} v_2^\top \Sigma_{22} w_2 \\ \rho_{13} v_3^\top \Sigma_{33} w_3 \end{bmatrix} \end{aligned} \quad (25)$$

where the last quadratic form does not involve v_1 , and we have with the usual argument that genvar produces w also. See Kettenring (1971) for

⁷See chapter two of T. K. Dijkstra (1981). *Latent variables in linear stochastic models* (PhD thesis). Second edition 1985, Sociometric Research Foundation, Amsterdam.

⁸R. G. D. Steel (1951). Minimum generalized variance for a set of linear functions. *Annals of Mathematical Statistics*, 22, 456-460.

an appropriate iterative routine (this involves the calculation of ordinary canonical variables of x_i and the $(N - 1)$ -vector consisting of the composites of the other blocks).

Another program is ‘maxvar’, which maximizes the largest eigenvalue. For every v one can calculate the linear combination of the corresponding composites that best predicts or explains them: the first principal component of $\Psi(v)$. No other set is as well explained by the first principal component as the maxvar composites. There is an explicit solution here, no iterative routine is needed, if one views the calculation of eigenvectors as non-iterative, see Kettenring (1971) for details⁹. One can show again that the optimal v equals w , although this requires a bit more work than for genvar (due to the additional detail needed to describe the solution).

As one may have expected, there is also ‘minvar’, the program aimed at minimizing the smallest eigenvalue (Kettenring (1971)). The result is a set of composites with the property that no other set is ‘as close to linear dependency’ as the minvar set. We also have an explicit solution, and w is optimal again.

As indicated in the introduction, PLS’ mode A can be used to retrieve w as well. Start with any v , and regress x_1 on the sign-weighted sum (instead of the other way around, as with mode B). This yields a vector proportional to

$$\text{sgn}(\Psi_{12}) \cdot \Sigma_{12}v_2 + \text{sgn}(\Psi_{13}) \cdot \Sigma_{13}v_3. \quad (26)$$

Inserting the expressions for Σ_{12} and Σ_{13} in terms of w yields quickly that the new update for v_1 is proportional to $\Sigma_{11}w_1$. The other mode A regressions yield vectors proportional to $\Sigma_{22}w_1$ and $\Sigma_{33}w_3$ respectively. So we have a fixed point of the PLS mode A equations, that can be trivially transformed into w .

A remark about the PLS-algorithms in this section: they were applied to path diagrams with links between *all* composites, whereas many path diagrams in practice will have at least some composites which are not connected directly. As long as the Σ_{ij} have the rank one product structure, that is irrelevant when we work with Σ . With S it is an empirical matter: it may be, as seems to be backed up by some experience, that leaving out composites in the sign-weighted sum that are only weakly connected with its corresponding composite helps reduce the sampling variability in the weight estimators.

A challenge, not for Σ but for S , is that one might have specified a simultaneous equation system for the composites, which leads to restrictions on their

⁹This is true when applied to S as well: methods other than maxvar and minvar will for S require more than just one iteration (and all programs produce different results).

correlations matrix. We could in principle employ them in the determination of the weights, and so enhance the statistical stability. The numerical issues remain to be resolved for arbitrary setups¹⁰.

3. Testing the composites factor model

Here we sketch four more or less related approaches to test the appropriateness or usefulness of the model. In practice one might perhaps want to deploy all of them.

3.1 Testing rank restrictions on sub-matrices using standard two-block canonical variables.

The covariance matrix of any sub-vector of x_i with any choice from the other indicators has rank one. Therefore one could use any of the methods developed for restricted rank testing, using standard canonical variables. A possible objection could be that the tests are probably sensitive to deviations from the Gaussian distribution, but jackknifing or bootstrapping might help to alleviate this. Another issue is the fact that we get many tests that are also correlated, so that simultaneous testing techniques based on Bonferroni, or more modern approaches are required¹¹.

3.2 Exploiting the difference between different estimators.

We noted that a number of generalized canonical variable programs yield identical results when applied to a Σ satisfying the composites factor model. But we expect to get different results when this is not the case. So, when using S one might want to check whether the differences between, say PLS and maxvar (or any other couple of methods), are too big for comfort. The scale on which to measure this could be based on the probability (as estimated by the bootstrap) of obtaining a larger ‘difference’ than actually observed¹².

3.3 Prediction tests, via cross-validation.

The path diagram might naturally indicate composites that are most relevant for prediction, as in the context of Technology Acceptance Models¹³ e.g. So

¹⁰See also T. K. Dijkstra (2010). Latent Variables and Indices: Herman Wold’s Basic Design and Partial Least Squares. Chapter 1 in V. Esposito Vinzi, W. W. Chin. J. Henseler, H. Wang (eds.): *Handbook of Partial Least Squares*. Springer.

¹¹See e.g. chapter 34 from A. DasGupta (2008). *Asymptotic Theory of Statistics and Probability*. Springer.

¹²It is not clear, to me, whether bootstrap samples should be generated from the original sample data or from ‘model transformed’ sample data (which means that the covariance matrix of the transformed data satisfies the composites factor model).

¹³See for this model F. D. Davis (1989). Perceived Usefulness, Perceived Ease of Use, and

it would seem to make sense to test whether the model's rank restrictions can help improve predictions of certain selected composites. As noted before, the result will not only reflect model adequacy but also the statistical phenomenon that the imposition of structure, even when strictly unwarranted, can help in prediction. So it would reflect also the sample size.

3.4 Global goodness-of-fit-tests.

In SEM we test the model by assessing the probability value of a Chi²-type distance measure between the sample covariance matrix S and a matrix $\widehat{\Sigma}$ that satisfies the model. Popular measures are $\frac{1}{2}\text{tr}\left(S^{-1}\left(S - \widehat{\Sigma}\right)\right)^2$, and $\text{tr}\left(S\widehat{\Sigma}^{-1}\right) + \log\left(\det\left(S^{-1}\widehat{\Sigma}\right)\right)$ minus the number of indicators. They belong to a large class of distances, all expressible in terms of a suitable function f :

$$\sum_k f\left(\gamma_k\left(S^{-1}\widehat{\Sigma}\right)\right). \quad (27)$$

Here $\gamma_k\left(S^{-1}\widehat{\Sigma}\right)$ is the k^{th} eigenvalue of its argument, and f is essentially a smooth real function defined on positive real numbers, with a unique global minimum of zero at the argument value 1. For the examples referred to we have $f(\gamma) = \frac{1}{2}(1 - \gamma)^2$ and $f(\gamma) = \gamma - \log(\gamma) - 1$ respectively. Another example is $f(\gamma) = \frac{1}{2}(\log(\gamma))^2$. The idea is that when the model fits perfectly $S^{-1}\widehat{\Sigma}$ is the identity matrix and all its eigenvalues equal one, and conversely. This class of distances was first analyzed by Swain¹⁴ (1975). Distance measures outside of this class are those induced by WLS with general fourth-order moments based weight matrices, but also the simple $\text{tr}\left(S - \widehat{\Sigma}\right)^2$. We can take any of these measures, calculate its value and use the bootstrap to estimate the corresponding probability value. A probably highly redundant reminder: the observation vectors ought to be pre-multiplied by $\widehat{\Sigma}^{\frac{1}{2}}S^{-\frac{1}{2}}$ (to ensure that their empirical distribution has a covariance matrix that agrees with the model) before the bootstrap is implemented. For $\widehat{\Sigma}$ one could take $\widehat{\Sigma}_{ii} := S_{ii}$ and for $i \neq j$

$$\widehat{\Sigma}_{ij} := \widehat{w}_i^{\text{T}} S_{ij} \widehat{w}_j \cdot S_{ii} \widehat{w}_i \cdot \widehat{w}_j^{\text{T}} S_{jj}. \quad (28)$$

4. Conclusion

User Acceptance of Information Technology. *MIS Quarterly*, 13, 319-340.

¹⁴A. J. Swain (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. *Psychometrika*, 40, 315-335. See also T. K. Dijkstra (1990). Some properties of estimated scale invariant covariance structures. *Psychometrika*, 55, 327-336.

Although the discussion of various generalized canonical variables programs could have been misleading for general readers, those readers familiar with PLS will not have failed to note what the paper is really all about: the specification of a model, in the classical sense of a restriction on (aspects of) the distribution function of the observables, that fits the PLS-algorithms and the underlying modeling approach as ‘naturally as possible’. The traditional way to expound PLS is to introduce a particular factor model with unobservable latent variables, the so-called ‘basic design’, and a set of alternating least squares algorithms that produce composites, who are taken as ‘proxies’ for the latent variables¹⁵. The relationships between the former represent an inevitable distortion of the relationships between the latter (consistency-at-large notwithstanding), and there is an inevitable stream of papers bemoaning this (the author’s PhD thesis included). There are also constructive contributions (the author’s PhD thesis and some recent work included) that correct the ‘inconsistencies’ in a simple way, and an additional ‘simple’ adaptation gives asymptotically efficient estimators on a par with ML¹⁶. But they all take as their point of departure the latent variables factor model, where PLS is by necessity an approximation. In the present paper we turn the tables as it were and define a composites factor model with observables only: now the latent variables factor models are approximations, and a perfect fit is generally not possible (but no doubt there is also a consistency-at-large property).

Spaß beiseite, the present paper should not be construed as an attempt to heat up a discussion that seems to be getting ever more personal¹⁷. An instrumentalist approach seems to be called for, aimed at testing the *usefulness* of the alternative approaches, and cataloguing with an open mind in what fields, and in which stages of development, and for which types of data they are most appropriate. Personally I would emphasize ‘prediction’ but I am willing to concede that understanding has some uses as well.

The ideas in this paper certainly need further analysis and development. A case in point is the incorporation in a user friendly way of the information embodied in simultaneous equation systems in the determination of the weights. Another, perhaps even more pressing challenge is to develop higher order stages, as in principal components and canonical variables analysis

¹⁵See H. Wold (1982). Soft Modeling: The Basic Design and Some Extensions. Chapter 1 in K. G. Jöreskog & H. Wold (1982), eds. *Systems under indirect observation, part II*. North-Holland, Amsterdam.

¹⁶See Wenjing Huang (2013). *PLSe: Efficient Estimators and Tests for Partial Least Squares*. PhD thesis UCLA. A paper version, together with P. Bentler and the present author is in preparation.

¹⁷References not included.

proper. The approach in this paper would be the first stage. For the next stage one would dig deeper, and extract path diagrams of other dimensions of the concepts under scrutiny. It would seem to be important to specify the model in such a way that the algorithms as described when applied to S return the first stage, even when there are more ‘layers of meaning’, instead of some mixture of composites at various dimensions. It is not evident, to the present author, whether and how this can be done.

*Acknowledgement.*¹⁸

¹⁸A recent review of a paper stimulated me to write this note. Proper acknowledgement will follow when that paper is published.