Communications of the Association for Information Systems

Volume 52 Paper 30

6-19-2023

Needed Developments in the Understanding of Quasi-Factor Methods

Edward E. Rigdon
Georgia State University, erigdon@gsu.edu

Communications of the Association for Information Systems

Debate Article DOI: 10.17705/1CAIS.05230 ISSN: 1529-3181

Needed Developments in the Understanding of Quasi-Factor Methods

Edward E. Rigdon

Marketing Department
Georgia State University
erigdon@gsu.edu

Abstract:

In this response to Evermann and Rönkkö (2023), I acknowledge areas of agreement about applying partial least squares (PLS) path modeling but cite substantial disagreements. The authors encourage researchers to also consider generalized structured component analysis and regression component analysis as procedures built around weighted composites rather than common factors. We can best regard all of these methods as quasi-factor methods, which avoid the uncertainty of factor indeterminacy though they still rely on covariance structures similar to those in factor analysis. They remain useful tools for researchers.

Keywords: Structural Equation Modeling, Partial Least Squares Path Modeling, Generalized Structured Component Analysis.

This manuscript underwent editorial review. It was received in 2021. John Venable served as Associate Editor.

1 Introduction

It is an honor to be invited to comment on Evermann and Rönkkö (2023). I certainly agree with some of the ideas expressed there. Rigdon (2016) flatly rejected the idea that partial least squares (PLS) path modeling somehow compensates for an otherwise inadequate sample size. Research should reduce uncertainty (Rigdon et al., 2020; Rigdon & Sarstedt, 2021), and inadequate sample size undermines that objective regardless of statistical method. While uncertainty in measurement arises from many other sources beyond random sampling variance, nevertheless, random sampling variance often represents a major contributor, which researchers can effectively control by maintaining an adequate sample size. Researchers need to focus on reducing uncertainty rather than on achieving "statistical significance", an essentially meaningless concept (Wasserstein et al., 2019). (In this regard, I find it disappointing that Evermann and Rönkkö (2023) embrace the outdated notion that "statistically significant" results qualitatively differ from results that are "not statistically significant" (Gelman & Stern 2006).

I also agree with the notion that using unweighted composites could be better than using a weighted composite method such as PLS path modeling—in some situations and in terms of some criteria (Dana & Dawes, 2004). Low sample size and low population R^2 can lead to both correlation weights and simple unit weights that outperform regression weights in out-of-sample prediction, a finding that extends to PLS path modeling (Becker et al., 2013).

The same goes for an appeal to "exploratory research" as an argument for choosing PLS path modeling (Rigdon, 2016). PLS path modeling, similar to common factor-based structural equation modeling (SEM), is a tool for making inferences about unobserved conceptual variables such as "perceived ease of use". "Exploratory" and "unobservable" hardly go together—it takes context and prior knowledge to weed out alternative explanations and validate the empirical proxies—whether common factors or weighted composites—that one uses to represent those unobservables in a statistical model. Calling work "exploratory" essentially amounts to a confession that it makes a minimal contribution to reducing uncertainty except in extreme cases.

I endorse Evermann and Rönkkö's (2022) dismissal of "polemic" papers. One dictionary defines "polemic" as "a strong verbal or written attack on someone or something" (OxfordLanguages 2022). With profoundly important issues, where researchers have invested much of themselves, it can be difficult to maintain scientific objectivity—to keep the focus on finding truth rather than on winning debating points. Polemic writings offer no value, though they can confuse and mislead the unwary.

In addition, I agree that generalized structured component analysis (GSCA) (Hwang & Takane, 2004, 2015) needs more consideration from researchers who consider PLS path modeling. Today, researchers will find a strong statistical basis, decent software, and interesting generalizations of the method to more complex research contexts such as the data that results from functional magnetic resonance imaging (Sarstedt & Hwang 2020). Moreover, GSCA results will necessarily mirror PLS path modeling and, thus, exhibit equivalent results in idealized situations and comparable results in simulations (Cho & Choi, 2020). I have been pleased to see the emerging collaboration between the GSCA / PLS path modeling camps as a recent special issue in *Behaviormetrika* reflects (Sarstedt & Hwang, 2020), and I look forward to more.

Some reluctance to take up GSCA may stem from a lack of trustworthy software. Indeed, a software error was one of the fatal flaws that invalidated Hwang et al.'s (2010) conclusions (Henseler, 2012; Hwang & Takane, 2015). Now, however, potential users have a range of strong software options, which includes packages in R such as cSEM (Rademaker & Schuberth, 2021) and standalone software programs such as GSCA Pro (Hwang et al., 2021). I have used both of these packages, and they work well.

2 Estimate Factor Model Parameters Using Composite-based Methods?

Evermann and Rönkkö (2023) recommend that, if researchers want to estimate factor model parameters using PLS path modeling (and one might extend that to GSCA), they should apply adjustments to produce consistent estimates of factor model parameters. Pointedly, Evermann and Rönkkö (2023) do not endorse using PLS path modeling to estimate factor model parameters and neither do I. I recall the late Theo Dijkstra saying that neither did Herman Wold, who created PLS path modeling. Sometimes, indeed, one cannot take the direct route to a destination. Browne (1979) first developed maximum likelihood estimators for the parameters of Tucker's (1958) interbattery factor model, which defied direct maximum likelihood estimation, using parameter estimates from a (composite-based) canonical correlation analysis as a stepping stone. However, one might also estimate such a recalcitrant factor model in an alternating fashion—estimate only a statistically identified subset of parameters while leaving the others at fixed values, and then turn the new estimates into fixed values and estimate the other set, alternating estimation of the two sets until convergence. This process need not involve too much complexity in programming. And it is what both PLS path modeling and GSCA do when estimating a composite model. So modelers could gain flexibility without switching between factor model and composite model.

3 Regression Component Analysis

I further endorse Evermann and Rönkkö's (2022) call for researchers to more explicitly present their methods. Of course, this issue represents an old, old quandary. When Karl Jöreskog and Dag Sörbom shifted the emphasis of their common factor package, LISREL, from matrix-based input to the SIMPLIS natural language syntax (Jöreskog & Sörbom 1993), observers worried that users would lose touch with the models they actually estimated. More recently, the authors of the amazingly flexible OpenMx package (Boker et al., 2021) philosophically opposed defaults of any kind and preferred that users explicitly specify every single element of their model—and yet found themselves creating application packages with defaults for certain kinds of applications in order to encourage adoption. Nevertheless, researchers need to understand the models they estimate or criticize and also need to explicitly report their research.

Schönemann and Steiger's (1976) regression component analysis (RCA) constitutes a clear and explicit composite-based approach to structural equation modeling. (Here, I borrow some points from a working paper.) Schönemann and Steiger (1976) developed RCA to avoid modeling with indeterminate common factors—statistical fabrications whose case-by-case values are intrinsically unknowable—when one might just as well use weighted sums of the model's observed variables. Those authors, who wrote extensively on factor indeterminacy (see also Guttman, 1955; Rigdon et al., 2019) could see little point in forcing researchers to deal with factor indeterminacy's uncertainity when they could so easily avoid it.

One begins with a common factor model:

$$\underline{y} = \Lambda \underline{\eta} + \underline{\varepsilon} \tag{1}$$

where $\underline{y}$ is a P x 1 vector of mean-centered observed variables, Λ is a P x K matrix of loadings, $\underline{\eta}$ is a K x 1 vector of common factors, and $\underline{\varepsilon}$ is a P x 1 vector of specific factors. Give both the common factors and the specific factors expected values of 0. Schönemann and Steiger's (1976, p. 177) approach replaces $\underline{\eta}$ and $\underline{\varepsilon}$ with weighted composites:

$$\underline{y} = \Lambda W' \underline{y} + \left(I - \Lambda W'\right) \underline{y} \tag{2}$$

where W is a P x K matrix of weights, the apostrophe ' indicates the matrix transpose, and I is an identity matrix of dimension P x P.

As Schönemann and Steiger (1976) pointed out, the loading matrix in Equation 2 could very well be the same loading matrix as in Equation 1. It could be exactly the same matrix, estimated in a factor model using maximum likelihood estimation, or any other estimator.

We can calculate the weight matrix W as (Schönemann & Steiger, 1976, p. 179):

$$W = \Sigma^{-1} \Lambda \left(\Lambda' \Sigma^{-1} \Lambda\right)^{-1} \tag{3}$$

where Σ is the covariance matrix of the observed variables, $^{-1}$ indicates the matrix inverse, and the apostrophe ' indicates the matrix transpose. Moreover, one could easily transform Λ and W', such as by using a square matrix T of dimension K by K, post-multiplying ΛT, and pre-multiplying T^{-1}W', without changing anything fundamental about the model.

GSCA, like PLS path modeling, offers users the ability to choose whether or not to include loadings. Without loadings, GSCA and PLS path modeling directly generalize canonical correlation analysis to more than two sets of variables. With loadings, we can represent the GSCA model as (Hwang & Takane, 2015, p. 19):

$$\underline{y} = \Lambda_G W_G' \underline{y} + \underline{e}_G \tag{4}$$

with the subscript G indicating GSCA. Algebra shows that:

$$\underline{e}_G = (I - \Lambda_G W_G')\underline{y} \tag{5}$$

so that one can represent the GSCA model as:

$$\underline{y} = \Lambda_G W_G' \underline{y} + \left(I - \Lambda_G W_G'\right)\underline{y} \tag{6}$$

Equations 6 and 2 show that the GSCA model with loadings closely resembles the RCA model except that the GSCA error terms incorporate the loading and weight matrices rather than representing them separately. Besides the impact of different estimation methods, the parameter estimates will differ because GSCA imposes a standardization constraint on the weighted composites W'$\underline{y}$—a necessary identification constraint for GSCA but not for RCA. But we can use the K x K covariance matrix of the weighted composites to standardize the RCA composites and, thus, obtain GSCA parameter estimates (under ideal conditions).

Relative to this level of clarity, PLS path modeling expositions tend to be vague except for works such as Tenenhaus et al. (2017). More explicit specification would be a very good thing for PLS path modeling.

3.1 Common Factors or Composites?

Are composite-based models "better" or "worse" than common factor models? In what sense? Data that can be described by a common factor model can also be described by composite-based RCA (Schönemann & Steiger, 1976)—if one needs only describe the data. Furthermore, one can construct an inferential test for the RCA model just as well as for the common factor model (Schönemann & Steiger, 1976, p. 184), but such a test would have the same practical limitations as factor-based SEM's χ^2 test (Jöreskog, 1969, p. 200).

Alternatively, if researchers took a scientific realist view (Rigdon, 2016, 2022) and saw their common factors or composites as proxies for unobserved conceptual variables, how could they know whether common factors or composites would be more valid—in any one instance, much less as a general rule? Evermann and Rönkkö's (2022) disdain for existing "validity metrics" does not go far enough. Not one extant metric actually quantifies the similarity between empirical proxy and unobserved conceptual variable, so none actually assess validity. They are holdovers from structural equation modeling's empiricist past, and we need to replace them.

3.2 Composite-Based Methods vs. Quasi-Factor Methods

Rigdon (2012) called for a clean break between factor-based SEM and PLS as a composite-based method. Rigdon (2012) wrongly promoted the label "composite-based methods". While these methods—RCA, PLS path modeling and GSCA—do model relations between composites, one can better think of them as quasi-factor methods. RCA, in particular, starts with a factor model. The way in which Cho and Choi (2020) describes an updated method for constructing data sets consistent with the GSCA model relies on factor model-like covariance structures. The label "quasi-factor methods" better captures this relationship. But that does not change the question in the prior section—which approach is better: the factor model with indeterminate common and specific factors, or the quasi-factor approach with weighted composites? We will not settle this decades-old question here (e.g., Velicer & Jackson 1990).

On a separate point, I find it disappointing that Evermann and Rönkkö (2023) seek to defend the misleading observations in Rönkkö and Evermann (2013) that lay behind their original dismissal of PLS path modeling. Rönkkö and Evermann (2013) specified a particular PLS path model that behaved bizarrely precisely because that model violated known conditions for PLS path modeling to perform well. One can do the same for common factor models. For example, imagine a common factor model with one common factor and any number of indicators. To achieve statistical identification, standardize the common factor to a variance of 1. A simulation would see the signs of the loadings flipping from positive to negative and back, from replication to replication, which would produce a bimodal and clearly non-normal distribution because, in such a model, the estimated loadings are indeterminate with regard to sign. Or consider a model with two common factors with two indicators each. Simulate data from a population where the true factor correlation is 0 as Rönkkö and Evermann (2013) did. Factor model estimation will fail because the model is not statistically identified (there is no unique best set of parameter estimates) at the population level, though random sampling variance may enable estimation by randomly producing factor correlations far enough from zero—in either direction, randomly. Rather than overgeneralizing and rejecting an entire methodology, it makes more sense simply to acknowledge and work within each method's limitations.

Evermann and Rönkkö (2023) claim that Rönkkö and Evermann (2013) used a "basic" model or a model that somehow typified PLS applications in an exploratory context. Exploratory analyses that involve empirical proxies for unobserved conceptual variables (as I mention above) almost certainly constitute a waste of time (Rigdon, 2016). In order to have any chance of validating those proxies and any inferences that derive from them, researchers must have substantial prior knowledge. An expansive nomological network needs to exist, and researchers must bring that knowledge to bear in assessing validity. Knowledge grows incrementally as researchers add a bit here and there to what we already know. Rönkkö and Evermann (2013) provided a poor exemplar model and an insufficient basis for the anathema that they pronounced.

4 Conclusion

One has room to appreciate regression component analysis, generalized structured component analysis, and PLS path modeling as valuable analytical methods. Not being common factor methods does not make them flawed—rather, it enables them to avoid an important shortcoming of the factor analytic approach. The dominance of common factor-based SEM can make it hard to appreciate the merits, and even the exact nature, of alternative procedures, but effective researchers will keep an open mind.

References

Becker, J.-M., Rai, A., & Rigdon, E. E. (2013). Predictive validity and formative measurement in structural equation modeling: Embracing practical relevance. In *Proceedings of the International Conference on Information Systems.*

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., Mehta, P., von Oertzen, T., Gore, R. J., Hunter, M. D., Hackett, D. C., Karch, J., Brandmaier, A., Pritikin, J. N., Kirkpatrick, R., Zahery, M. (2021). *OpenMx user guide*. Virginia Commonwealth University.

Browne, M. W. (1979). The maximum-likelihood solution in inter-battery factor analysis. *British Journal of Mathematical and Statistical Psychology, 32*(1), 75-86.

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.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, *29*(3), 317-331.

Evermann, J., & Rönkkö, M. (2023). Recent developments in PLS. *Communications of the Association for Information Systems, 52*.

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, *60*(4), 328-331.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology, 8*(2), 65-81.

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.

Hwang, H., Cho, G., & Choo, H. (2021). *GSCA Pro 1.1*. Retrieved from https://www.gscapro.com/

Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., & Hong, S. (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research*, *47*(4), 699-712.

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81-99.

Hwang, H., & Takane, Y. (2015). *Generalized structured component analysis: A component-based approach to structural equation modeling*. CRC Press.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183-202.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific Software International.

OxfordLanguages (2022). *Oxford English Dictionary*. Oxford University.

Rademaker, M., & Schuberth, F. (2021). *cSEM: Composite-based structural equation modeling*. CRAN. Retrieved from https://CRAN.R-project.org/package=cSEM

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5-6), 341-358.

Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal*, *34*(6), 598-605.

Rigdon, E. E. (2022). The proxy of Dorian Gray: Scientific realism, construct validation, and the way forward. *European Journal of Marketing, 57*(6), 1725-1736.

Rigdon, E. E., Becker, J. M., & Sarstedt, M. (2019). Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. *Multivariate Behavioral Research*, *54*(3), 429-443.

Rigdon, E. E., & Sarstedt, M. (2021). Accounting for uncertainty in the measurement of unobservable marketing phenomena. *Review of Marketing Research*. In H. Baumgartner & B. Weijters (Eds.), Review of marketing research. Emerald.

Rigdon, E. E., Sarstedt, M., & Becker, J. M. (2020). Quantify uncertainty in behavioral research. *Nature Human Behaviour*, *4*(4), 329-331.

Sarstedt, M., & Hwang, H. (2020). Advances in composite-based structural equation modeling. *Behaviormetrika, 47*, 213-217.

Schönemann, P. H., & Steiger, J. H. (1976). Regression component analysis. *British Journal of Mathematical and Statistical Psychology*, *29*(2), 175-189.

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. (2017). Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. *Psychometrika*, *82*(3), 737-777.

Tucker, L. R. (1958). An inter-battery method of factor analysis. *Psychometrika, 23*(2), 111-136.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*(1), 1-28.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p < 0.05”. *The American Statistician*, *73*(1), 1-19.

.

About the Author

Edward Rigdon is professor of marketing in the J. Mack Robinson College of Business at Georgia State University in Atlanta, Georgia, USA. He has been working with structural equation modeling (SEM) methods since the late 1980s. His research has appeared in a variety of high quality business journals, including *MIS Quarterly* and *Information Systems Research*, as well as in quantitative methods journals including *Psychometrika* and *Multivariate Behavioral Research*. In 1993, he co-founded semnet, an email discussion list devoted to SEM, which currently has more than 3,500 subscribers from more than 50 countries. He currently focuses on accounting for uncertainty in social science research.