Technological Forecasting & Social Change 194 (2023) 122665

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

# The choice of structural equation modeling technique matters: A commentary on Dash and Paul (2021)☆

Florian Schuberth[a,*], Geoffrey Hubona[b], Ellen Roemer[c], Sam Zaza[d], Tamara Schamberger[e,a], Francis Chuah[f], Gabriel Cepeda-Carrión[g], Jörg Henseler[a,h]

[a] Faculty of Engineering Technology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
[b] Texas A&M International University, 5201 University Boulevard, Laredo, TX 78041, USA
[c] Department of Business Administration and Economics, Ruhr West University of Applied Sciences, Duisburger Str. 100, 45479 Mülheim an der Ruhr, Germany
[d] Information Systems and Analytics Department, Jones College of Business, Middle Tennessee State University, Murfreesboro, TN 37132, USA
[e] Faculty of Business Management and Economics, University of Würzburg, Sanderring 2, 97070, Würzburg, Germany
[f] School of Business Management, Universiti Utara Malaysia, Sintok, Kedah, Malaysia
[g] Department of Business Management & Marketing, Universidad de Sevilla, Av. Ramón y Cajal, 1, 41018 Sevilla, Spain
[h] Nova Information Management School, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal

A R T I C L E  I N F O

Dataset link: https://osf.io/z2h7v/?view_only=52e0d1e574dc42e2a22b92192d87c16a

*Keywords:*
Structural equation modeling (SEM)
Covariance-based structural equation modeling (CB-SEM)
Partial least squares based structural equation modeling (PLS-SEM)
Consistent partial least squares (PLSc)

A B S T R A C T

Ganesh Dash and Justin Paul authored an article titled “CB-SEM vs. PLS-SEM methods for research in social science and technological forecasting” in a special issue of Technological Forecasting and Social Change, co-edited by Justin Paul. Unfortunately, the article’s central conclusion – “CB or PLS or PLSc do not matter” – is misleading and at odds with practically all extant conceptual and empirical research on this subject. This commentary identifies an unsuitable research design to be the major cause of the erroneous conclusion and aims to set the record straight. A Monte Carlo simulation demonstrates that the choice of the approach to structural equation modeling can have a substantial impact on the results and their validity. In general, analysts should choose a structural equation modeling approach that fits their conceptual model.

## 1. Motivation

Research in technological forecasting and social change seeks to explain and predict the relationships between social, environmental, and technological factors. To this end, researchers use various statistical methods, including structural equation modeling (SEM). Since different approaches to SEM coexist, it is important to understand their (relative) performances in guiding analysts who use them in empirical research. An article that aims to contribute to this endeavor is one by Dash and Paul (2021), recently published in this journal, in which the authors analyze the relative performances of the three approaches to SEM, i.e., covariance-based SEM (CB-SEM), partial least squares SEM (PLS-SEM), and consistent PLS (PLSc), using an empirical example.

The central conclusion at which Dash and Paul (2021, p. 9) arrive is “[t]here is no single way to excellence. CB or PLS or PLSc do not matter”. This conclusion is surprising, because it is at odds with practically all extant research on that topic. Other studies have found differences between CB-SEM, PLS-SEM, and PLSc with regard to the consistency of parameter estimates (c.f. Dijkstra, 1981; Dijkstra and Henseler, 2015a,b; Hair et al., 2017b), the proneness to Type-I and Type-II errors (c.f. Goodhue et al., 2017), and the convergence behavior

☆ Acknowledgments: The last author gratefully acknowledges financial support from FCT Fundação para a Ciência e a Tecnologia (Portugal) national funding through a research grant from the Information Management Research Center – MagIC/NOVA IMS (UIDB/04152/2020). He also acknowledges a financial interest in the composite-based SEM software ADANCO and its distributor, Composite Modeling. The authors thank Alexandra Elbakyan for her efforts in making science accessible. Last but not least, the authors especially thank the editor of Technological Forecasting and Social Change, Scott Cunningham, for granting the opportunity to write this commentary.

* Corresponding author.
*E-mail addresses:* f.schuberth@utwente.nl (F. Schuberth), ghubona@gmail.com (G. Hubona), ellen.roemer@hs-ruhrwest.de (E. Roemer), sam.zaza@mtsu.edu (S. Zaza), tamara.schamberger@uni-wuerzburg.de (T. Schamberger), francischuah@uum.edu.my (F. Chuah), gabi@us.es (G. Cepeda-Carrión), j.henseler@utwente.nl (J. Henseler).

https://doi.org/10.1016/j.techfore.2023.122665
Received 16 February 2023; Received in revised form 11 May 2023; Accepted 21 May 2023
Available online 16 June 2023
0040-1625/

(c.f. Reinartz et al., 2009). The question then arises whether it really does not matter which approach to SEM is used.

By means of conceptual reasoning and a Monte Carlo simulation, this commentary demonstrates that the central conclusion of Dash and Paul (2021) is wrong – which SEM approach is chosen *does* matter indeed. In trying to determine the reasons why Dash and Paul (2021) reached such an erroneous conclusion, this commentary identifies an inappropriate research design as the primary factor, while problems with referencing and unclear or ambiguous language were contributing factors.[1]

## 2. CB-SEM, PLS-SEM or PLSc – Does it matter?

The central conclusion of Dash and Paul (2021, p. 9) can be stated as follows: "There is no single way to excellence. CB or PLS or PLSc do not matter". This conclusion is surprising in many regards. Not only is it at odds with their own example, in which they find that the different methods yield different results; it is also contrary to practically all methodological research comparing CB-SEM, PLS-SEM, and PLSc. In the following subsections, we show that the conclusion is wrong. We explain the conceptual differences between the three approaches based on the existing literature, demonstrate their different behaviors in a Monte Carlo simulation, and formulate recommendations for empirical researchers who want to make an informed choice between these structural equation modeling approaches.

### *Conceptual differences*

SEM is a flexible modeling approach that allows researchers to specify statistical models made up of equations explaining the relationships between constructs and their relationships with observed variables (e.g., Henseler, 2020). There are three dominant models to specify the relationship between a construct and its observed variables (e.g., Bollen and Diamantopoulos, 2017; Yu et al., 2021): the reflective measurement model also known as the common factor model (see Fig. 1(a)), the causal-formative measurement model (see Fig. 1(b)), and the composite model (see Fig. 1(c)). While in the reflective and causal-formative measurement model, the relationships between a *latent variable* (displayed as an oval) and observed variables (displayed as rectangles) are modeled, in the composite model, the relationships between an *emergent variable* (displayed as a hexagon) and observed variables are specified. While a latent variable is one "for which there is no sample realization for at least some observations in a given sample" (Bollen, 2002, p. 612), an emergent variable is a composite that conveys the information between its components and other variables of the model (e.g., Henseler and Schuberth, 2020b). For an elaborate exposition of the different models, their application, and for examples of the two different types of construct, we refer the interested reader to Bollen and Bauldry (2011), Henseler and Schuberth (2020a), Yu et al. (2021).

[1] One of the anonymous reviewers and the editor-in-chief, Scott Cunningham, asked us to disclose our personal motivation for writing this commentary, and we are happy to do so. When we became aware of the paper by Dash and Paul (2021), we were surprised by the number of apparently erroneous statements and references (see the Appendix for a selection). Consequently, the Dash and Paul (2021) paper makes the impression of a "lack of concern for the truth" (a term introduced by Meibauer, 2016, p. 71). While "[s]cience is self-correcting, in the sense that a falsehood injected into the body of scientific knowledge will eventually be discovered and rejected" (Goodstein, 2010, p. 2), such self-correction can take a long time, and newer research even conjectures that the self-correcting nature of science is nothing but a myth (Stroebe et al., 2012). Self-correction in science occurs primarily through the peer-review system and replication (Broad and Wade, 1982). We offer this commentary as a form of replication. We would like to help other researchers get a clear picture of the current state of SEM research and to have their own, fair judgement on the Dash and Paul (2021) paper. We declare that no member of the author team knows or has ever met Ganesh Dash or Justin Paul.

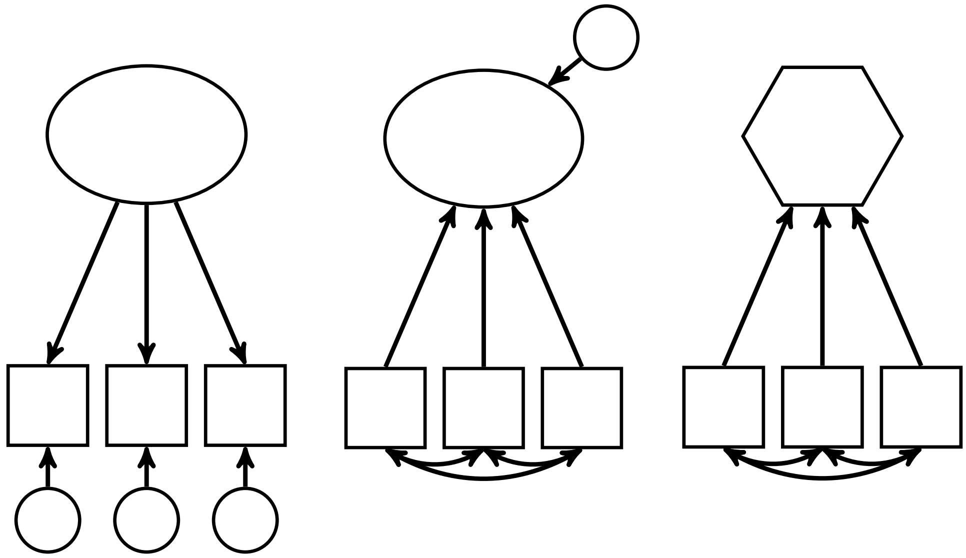
Once researchers have decided on the type of construct, i.e., a latent variable or an emergent variable, the relationships between the constructs and their observed variables, and the relationships between the constructs, they need to choose an approach for model parameter estimation. This is where CB-SEM, PLS-SEM, and PLSc come into play: They are all estimators for structural equation models (Evermann and Rönkkö, 2021). In general, an estimator should be chosen to fit the underlying statistical model as well as its assumptions. Otherwise, the estimator will most likely produce inconsistent and/or biased estimates which can jeopardize the conclusions drawn from an estimated model. While an unbiased estimator produces estimates with expected values equal to their population counterparts, a consistent estimator produces estimates that converge in probability to the population counterparts if the estimator's assumptions hold (e.g., Amemiya, 1985). As highlighted by Wooldridge (2012, p. 169) "Although not all useful estimators are unbiased, virtually all economists agree that consistency is a minimal requirement for an estimator. The Nobel Prize-winning econometrician Clive W. J. Granger once remarked, 'If you can't get it right as n goes to infinity, you shouldn't be in this business.' The implication is that, if your estimator of a particular population parameter is not consistent, then you are wasting your time".

CB-SEM, which actually refers to a class of estimators that include the maximum-likelihood estimator (Jöreskog, 1970) and the weighted least squares estimator, also known as asymptotic distribution free estimator (Browne, 1984), is arguably the most widely used approach for model parameter estimation. This approach minimizes a discrepancy between the sample variance–covariance matrix and the model-implied counterpart to obtain the parameter estimates. As has frequently been discussed in the literature, if the model is correctly specified, it provides consistent estimates for structural equation models containing latent variables (e.g., Bollen, 1989).

PLS-SEM, originally known as partial least squares path modeling, was proposed by Wold (1975) as a computationally efficient estimator to SEM. Hence, we agree with Dash and Paul's (2021, p. 9) remark that "PLS quickly estimates the cause-and-effect relationships (complex)". However, nowadays, with the current trend of increasing computer power, this advantage plays only a minor role. To estimate the model parameters, in a first step, PLS-SEM employs an iterative algorithm based on ordinary least squares regressions to determine weights for calculating construct scores (e.g., Lohmöller, 1989, Chapter 2). Subsequently, the construct scores are used to estimate the structural parameters. Against this background, it is not clear what is meant by "the PLS approach has become quite popular among researchers due to its variance based relationship rather than covariance" (Dash and Paul, 2021, p. 1), "[u]nlike Amos, SmartPLS is a tool based on partial least squares rather than covariance" (p. 5), or "[CB-SEM] is based on covariance, and [PLS-SEM] is based on variance (partial least squares)" (Abstract). Since PLS-SEM is based on ordinary least squares regressions, it takes into account both variables' variances and covariances to estimate regression coefficients.

In the literature it is well documented that PLS-SEM is unsuitable for estimating models comprising latent variables and, therefore, also unsuitable for estimating the parameters of reflective and causal-formative measurement models (Sarstedt et al., 2016). In such cases, PLS-SEM produces inconsistent estimates (e.g., Dijkstra, 1981; Hui and Wold, 1982) and, consequently, researchers are likely to draw questionable conclusions from their estimated models.

This also casts doubt on Dash and Paul's (2021, p. 8) claim that "[t]he most significant advantage of PLS-SEM is that both formative and reflective measurement models can be specified with it, whereas CB-SEM is limited to reflective models only". In fact, PLS-SEM's factor loadings estimates are typically biased upwards, while the estimated correlations between the latent variables are biased downwards due to attenuation (Gefen et al., 2011). This can also explain Dash and Paul's (2021) findings in their illustrative example, e.g., "the item loadings are usually higher in PLS-SEM than CB-SEM" (p. 6), "PLSc provides a



(a) Reflective measurement model (b) Causal-formative measurement model (c) Composite model

**Fig. 1.** Three different dominant models to specify the relationships between a construct and its observed variables. *Note:* An oval is used to represent the latent variables, while a hexagon represents the emergent variable.
*Source:* Adopted from Schuberth (2021a)

lower item loading compared to PLS and is closer to CB" (p. 7), and "[f]indings indicate that the item loadings are usually higher in PLS-SEM than CB-SEM" (Abstract). Similarly, it can explain why "PLS-SEM retains more measurement items under the constructs than CB-SEM" in Dash and Paul (2021, p. 8) when the decision about whether to retain an item is based on distorted PLS-SEM loading estimates.

PLS-SEM's parameter inconsistency also explains why the following conclusions about the average variance extracted (AVE) and composite reliability (CR) Dash and Paul (2021) draw are incorrect: "Hence, it can be concluded that the PLS method provides a more consistent item loading that boosts the reliability and validity of the factors" (p. 6), "It is also found that average variance extracted (AVE) and composite reliability (CR) values are higher in the PLS-SEM method, indicating better construct reliability and validity" (Abstract), and "However, the indicator reliability and validity go down in a consistent PLS algorithm compared to PLS" (p. 7). Metrics such as the AVE and the CR were developed against the backdrop of a reflective measurement model and are calculated based on the reflective measurement model's parameters (e.g., Fornell and Larcker, 1981; Jöreskog, 1971b). Consequently, these metrics typically show a higher value under PLS-SEM as Dash and Paul observed, e.g., "[a]lthough AVE is greater than 0.5 and CR is greater than 0.8, PLS-SEM produced higher values than CB-SEM in both methods" (p. 7). However, this is because AVE and CR are calculated based on inconsistent PLS-SEM parameters and not because PLS-SEM "boosts the reliability and validity of the factors" (Dash and Paul, 2021, p. 6). In fact, the use of metrics such as the AVE and CR in the PLS-SEM context has been criticized in the literature (e.g., Rönkkö and Evermann, 2013; McIntosh et al., 2014; Rönkkö et al., 2016; Schuberth, 2021a). Rönkkö et al. (2016) already recognized this, which also casts doubt on Dash and Paul's (2021) claim that "[one] school of thought completely disregard PLS" referring to Rönkkö et al. (2016). Notably, it was Rönkkö and colleagues who initiated a scientific debate that revealed various shortcomings of PLS-SEM (Rönkkö and Evermann, 2013). Finally, we recommend that researchers should assess the reliability of scores obtained by PLS-SEM Mode A weights by Dijkstra–Henseler's $\rho_A$ (Dijkstra and Henseler, 2015b), and not by the CR (Werts et al., 1974). The latter estimates the reliability of sum scores and not the reliability of construct scores obtained by Mode A weights.

To address PLS-SEM's inconsistency for models comprising reflectively measured latent variables, PLSc was introduced (Dijkstra and Schermelleh-Engel, 2014; Dijkstra and Henseler, 2015a,b). Similar to PLS-SEM, in a first step, PLSc determines weights by the iterative PLS algorithm using Mode A to calculate construct scores. However, and in contrast to PLS-SEM, in the second step, it is acknowledged that PLS-SEM's construct scores are contaminated by random measurement error in case of reflective measurement models. For this reason, it applies a correction for attenuation to obtain consistent estimates for the path coefficients among latent variables and factor loadings (Dijkstra and Henseler, 2015b). Hence, PLSc shows similar statistical properties as CB-SEM, i.e., both are consistent estimators for this type of model if the model is correctly specified. However, one has to note that PLSc is not suitable for all types of latent variable models that can be estimated by CB-SEM. For instance, while it is possible to estimate models containing correlations between random measurement errors within a block of indicators (Rademaker et al., 2019), it is currently not possible to estimate models containing correlations between random measurement errors in two different blocks as can be done by CB-SEM. For an elaborate exposition of the models that can be estimated by PLSc, we refer the interested reader to Schuberth et al. (2023). Against this background, we hope that we could rule out the confusion caused by the introduction of PLSc mentioned in Dash and Paul (2021, p. 4).

Considering structural equation models containing emergent variables, i.e., where all theoretical concepts have been operationalized by a composite model, it was shown that PLS-SEM provides consistent estimates (Dijkstra, 2017). In contrast, in CB-SEM it was for long not clear how to specify and thus estimate such models (see, e.g., Aguirre-Urreta and Marakas, 2014b; Rigdon et al., 2014; Aguirre-Urreta and Marakas, 2014a). However, the recently introduced Henseler–Ogasawara (H–O) specification addresses this issue and allows for a flexible way of modeling composites in CB-SEM (Henseler, 2020; Schuberth, 2021b).[2] Consequently, statements such as "[f]urther, to be specific, if the researchers' primary objective is to estimate a factor-based model, CB-SEM is the preferred one. On the other hand, if the

[2] We acknowledge that Dash and Paul (2021) may not yet have had access to these publications at the time of writing their paper.

primary aim is to estimate a composite-based model, PLS-SEM should be considered" (Dash and Paul, 2021, p. 8), and "CB-SEM does not support this as one is stuck with a factor-based approach" (Dash and Paul, 2021, p. 8), are outdated since, as explained, composite models can be estimated consistently by both PLS-SEM and CB-SEM.

Sometimes, in empirical research, scientists face structural models containing both latent and emergent variables (e.g., Benitez et al., 2020; Hwang et al., 2021). In this case, they can rely on both PLSc or CB-SEM. Obviously if PLSc is used, no correction is applied for constructs modeled as emergent variables (e.g., Schuberth et al., 2018b). We refer the interested reader to Table 1 in Schuberth et al. (2023) about the suitability of an estimator for a given model, i.e., for a reflective measurement model, a causal-formative measurement model, and a composite model.

Besides applying assessment criteria known from confirmatory and explanatory research, i.e., for theory testing (Henseler, 2018; Benitez et al., 2020), it was proposed that researchers evaluate the predictive power of models estimated by PLS-SEM (e.g., Shmueli et al., 2016). This is also recognized by Dash and Paul (2021, p. 8) when they state that "PLS-SEM can be used for prediction and explanation, whereas CB-SEM is limited to explanation (Hair Jr et al. 2017a; Wold, 1974)".[3] However, in their study Dash and Paul did not consider any metrics to judge the out-of-sample predictive power of their model. Hence, we assume that their research is grounded in explanatory rather than predictive research.

*Monte Carlo simulation*

To further highlight that it matters whether PLS-SEM, CB-SEM, or PLSc is used, we conducted a Monte Carlo simulation where we make use of two different population models. The two population models have the same structural model. As Fig. 2 shows, this is a full mediation model. The single difference between the populations is the way in which the constructs are defined. While in the first population model all constructs are defined as latent variables, in the second population model all constructs are defined as emergent variables. In both population models, the observed variables as well as the emergent and latent variables, respectively, have a mean of zero and a unit variance, i.e., all parameters are standardized parameters. Fig. 2 shows that for each construct the relationships with its observed variables are set to 0.6, 0.7, and 0.8. Note that in the latent variable population model, these values represent the values of the factor loadings, whereas in the emergent variable population model, these values represent the values of the composite loadings.

To study the approaches' performance, we considered five different sample sizes: 100 ($=10^2$), 316($\approx 10^{2.5}$), 1,000 ($=10^3$), 3,162 ($\approx 10^{3.5}$) and 10,000 ($=10^4$) observations. For each sample size and population model, we applied CB-SEM, PLS-SEM and PLSc. All calculations were carried out in the statistical environment R (version 4.2.3, R. Core Team, 2023). In doing so, we drew 1,000 samples from a multivariate normal distribution with mean zero and a variance–covariance matrix equal to the correlation matrix implied by each population model. To generate the data, we used the R package MASS (version 7.3-58.3, Venables and Ripley, 2002). In cases where one or more approaches did not converge or produced an inadmissible estimation for a specific sample, we discarded this sample, drew a new sample, and applied all approaches to this new sample. Consequently, we based the results of all approaches on the same 1,000 samples. In addition, we included a sample with a variance–covariance matrix that is identical to the one implied by the respective population model. This situation is denoted by $\infty$ in Figs. 3 and 4.

To estimate the population parameters, we employed CB-SEM, PLS-SEM, and PLSc using different model specifications and settings. Specifically, for CB-SEM, we used the maximum likelihood estimator as implemented in the R package lavaan (version 0.6-15, Rosseel, 2012). In doing so, we employed two model specifications, namely a factor model and a composite model specification. While in the factor model specification all constructs were specified as reflectively measured latent variables, as so-called common factors, as is typically done in CB-SEM, in the composite model specification the constructs were modeled as emergent variables, as proposed in the H–O specification (Schuberth, 2021b; Yu et al., 2023). Considering PLS-SEM, we employed two outer weighting schemes, i.e., Mode A and Mode B, as implemented in the R package cSEM (version 0.5.0, Rademaker and Schuberth, 2020). Similarly, for PLSc we employed Mode A in combination with a correction for attenuation, as proposed by Dijkstra and Henseler (2015b). In all model specifications the structural model was specified as in the two population models. The R code to conduct the analysis including the observed variables' population variance–covariance matrix can be downloaded from the following website: https://osf.io/z2h7v/?view_only=52e0d1e574dc42e2a22b92192d87c16a

Figs. 3 and 4 illustrate our simulation results for the emergent variable population and the latent variable population, respectively. Specifically, the figures show the average parameter estimates over the 1,000 simulation runs, including their 95% confidence intervals for the different sample sizes. The population value of a parameter is highlighted by the dashed line. Since the results for the composite loadings and factor loadings, respectively, were very similar, we illustrate only the results for the loadings of $x_1$, $m_2$, and $y_3$.

Fig. 3 displays the results for the emergent variable population. It shows that among the studied approaches only CB-SEM based on the composite model and PLS-SEM using Mode B produce estimates that, on average, are close to the population value. As expected for an increasing sample size, the estimates become more precise, as the tighter confidence intervals indicate. In addition, the two approaches could both retrieve the population parameters from the population variance–covariance matrix. In contrast, CB-SEM based on the factor model, PLS-SEM using Mode A, and PLSc produce biased estimates. This bias does not diminish for an increasing sample size.[4] Similarly, none of the three approaches were able to retrieve the population parameters from the population variance–covariance matrix of the observed variables. Consequently, a researcher could conclude that $\xi$ shows an effect on $\eta$, even though such an effect does not exist.

Fig. 4 shows the results for the latent variable population. As expected, CB-SEM based on the factor model and PLSc produce estimates that, on average, are very close to the true parameter values. Moreover, both approaches were able to retrieve the population parameters from the observed variables' population variance–covariance matrix. In contrast, CB-SEM based on the composite model, PLS-SEM using Mode B, and PLS-SEM using Mode A show a clear bias in their estimates for the latent variable population model. In particular, PLS-SEM shows upward biased factor loading estimates, as mentioned in earlier literature (Gefen et al., 2011). Further, a researcher could wrongly conclude that $\xi$ has an effect on $\eta$.

Overall, our results are in line with the findings of previous studies that compared CB-SEM, PLS-SEM, and PLSc (e.g., Rönkkö and Evermann, 2013; Dijkstra and Henseler, 2015a; Sarstedt et al., 2016; Schuberth, 2021a).

[3] In contrast to what is widely believed, out-of-sample predictions can also be obtained from a model estimated by CB-SEM (de Rooij et al., 2022). However, de Rooij et al.'s paper demonstrating this appeared after Dash and Paul (2021).

[4] Only recently, a second type of composite, the so-called *nomological composite* or *nomological component*, was introduced (Cho and Choi, 2020). In contrast to the classical composite discussed in our manuscript, the nomological composite is composed of its indicators using correlation weights. Models containing this type of composite can be estimated consistently by PLS Mode A. Considering that this type of composite is a special case of the classical composite presented in our manuscript (Cho et al., 2022), all approaches that consistently estimate models containing classical composites, also consistently estimate models containing this type of composite.

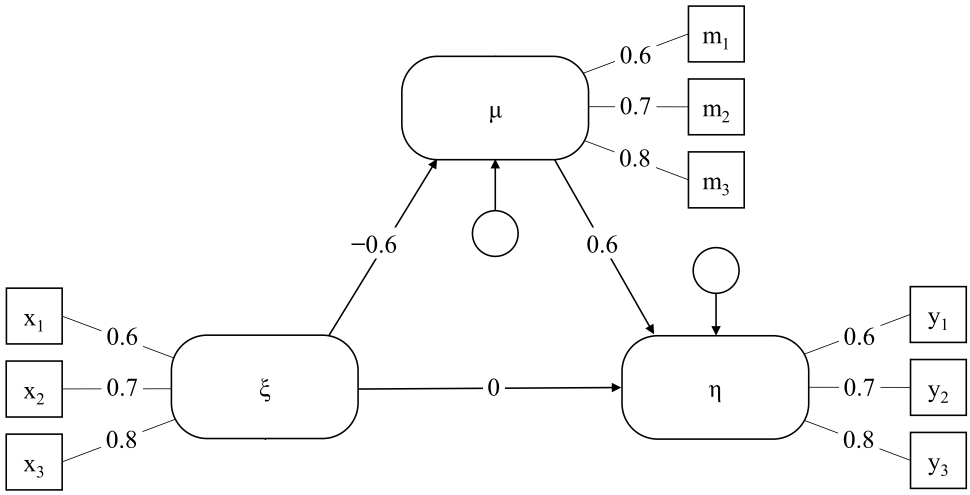



**Fig. 2.** Structure of the population model used in the Monte Carlo simulation.

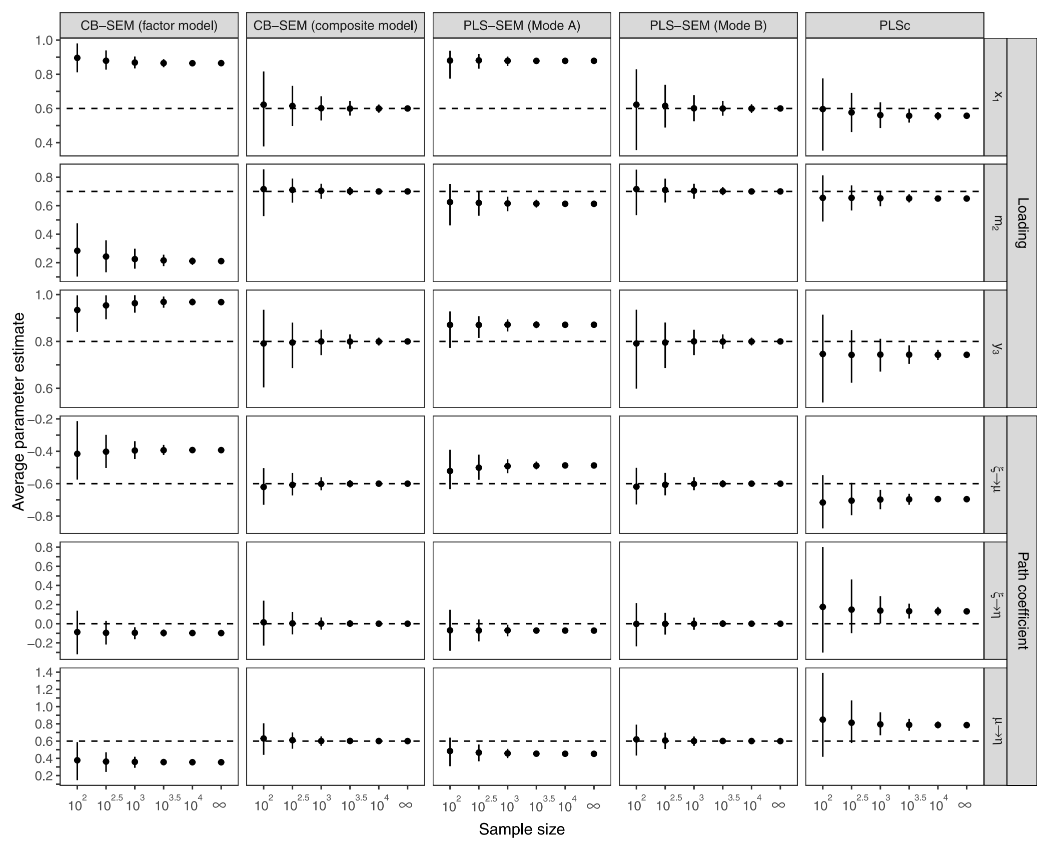



**Fig. 3.** Average parameter estimates and their 95% confidence intervals for the emergent variable population.

*Recommendation*

In the light of the conceptual differences between CB-SEM, PLS-SEM, and PLSc, as well as the results of various Monte Carlo simulations, including the one conducted in this study, we can thus confirm a very simple rule: In explanatory and confirmatory research, one should specify structural equation models in such a way that they correspond to the theoretical model, i.e., to the researchers' hypothesized working principle of the world. Only if the parameters of the structural equation model match the relationships hypothesized in the real world can the model be employed to make inferences about the unobservable underlying mechanism of the world. Consequently, it is important that researchers specify the relationships between their studied constructs in line with their proposed theory. Likewise, they should pay the same attention to the specification of their constructs. On the one hand, if an analyst studies emergent variables, then CB-SEM using the H–O specification or PLS-SEM using Mode B would be viable options. On the other hand, if an analyst studies reflectively measured latent variables, CB-SEM using a factor model specification or PLSc would be appropriate. Since some methods are limited to certain specifications, e.g., PLS-SEM (without correction for attenuation) is not suitable for models involving latent variables, we can conclude that the choice of method *does* matter.

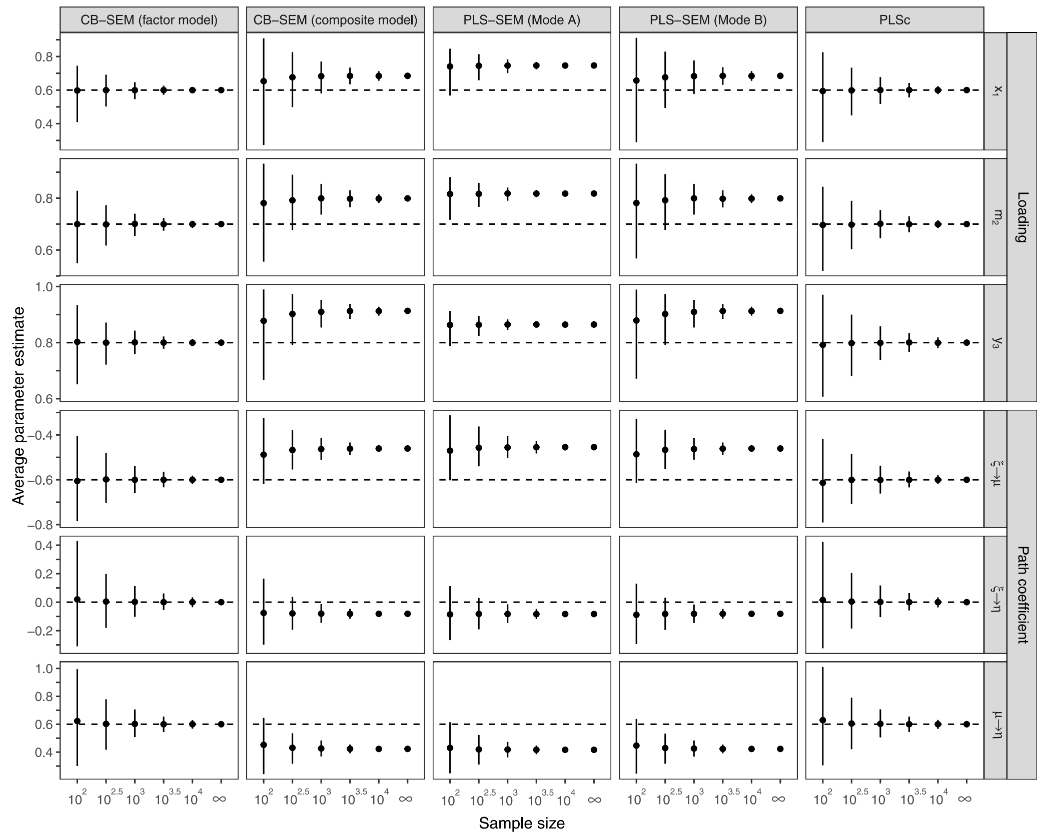



**Fig. 4.** Average parameter estimates and their 95% confidence intervals for the latent variable population.

## 3. What has led Dash and Paul (2021) to the erroneous conclusion?

In the previous section, we demonstrated that Dash and Paul's (2021) central conclusion – "CB or PLS or PLSc do not matter" – is erroneous and requires a more elaborate response. Therefore, we now ask how their study could have come to such a false conclusion. We have identified, and here discuss, an inappropriate research design as the primary factor, along with two contributing factors, i.e., (i) problematic referencing and (ii) unclear or ambiguous language.

*Primary factor: An inappropriate research design for comparing methods*

Research questions are the drivers of research designs, i.e., "[e]stablishing research questions also makes it possible to select research strategies and methods with confidence. In other words, a *research project is built on the foundation of its research questions*" (Blaikie, 2000, p. 58; emphasis in original). Thus, the research design is not an arbitrary choice; in order to arrive at plausible answers, it should be consistent with the research question(s) being investigated. Most importantly, "[d]ecisions made during the research design process ultimately impact the degree of confidence readers can place in the conclusions drawn from a study, the degree to which the results provide a strong test of the researcher's arguments, and the degree to which alternative explanations can be discounted" (Bono and McNamara, 2011, p. 656).

In their paper, Dash and Paul (2021, p. 2) aim to answer the following two research questions: "RQ1: which one is the better approach: CB-based model or a PLS-based model? For the same data (sample group), which one is better to be adopted (a piece of empirical evidence is provided)? RQ2: How do PLS and PLSc (consistent) differ, and which one is the actual counterpart of CB-SEM?"

First, Dash and Paul's (2021) first research question would require a comparison of the PLS-based and CB-based models. However, in their paper they do not explicitly indicate which statistical model(s) they are actually investigating. From their use of assessment criteria such as AVE and CR, which assume an underlying reflective measurement model, one could conclude that their statistical model comprises reflective measurement models and thus latent variables. This would be consistent with the stated limitation that "the specified model used in this study was not composite-based" (Dash and Paul, 2021, p. 8). However, because PLS-SEM consistently estimates the parameters of composite models (Hair et al., 2017b), not reflective measurement models, the employment of reflective measurement models would preclude the use of PLS-SEM.

Second, and even more importantly, Dash and Paul (2021) used a research design that cannot deliver an answer to their research questions. Specifically, they relied on an illustrative empirical example as the basis for comparing the relative performance of PLS-SEM, CB-SEM, and PLSc. However, analyzing empirical data using three different approaches does not allow us to draw valid conclusions about the effectiveness of these approaches. The study design has two unknowns: (1) the behavior of the approaches, and (2) the population model underlying the empirical data. Just as an equation with two unknowns cannot be solved uniquely, it is not possible to gain insight from an empirical study if neither the behavior of the research method nor the underlying population model are known. In fact, for empirical datasets, the data generating process, i.e., the population model which is responsible for the data at hand, is unknown (Goodhue et al., 2012).

A more promising alternative for comparing the performance of approaches would be to use Monte Carlo simulations, because they allow researchers to determine all the conditions under which an approach is studied, such as the distributions of the variables and the model structure in the population (see, e.g., Paxton et al., 2001; Schamberger, 2023). For Dash and Paul's (2021) study, this means that no valid conclusions can be drawn because any differences between the results of the statistical approaches could be due to random peculiarities of the particular dataset.

*Contributing factor 1: Problematic referencing*

Careful referencing of the work of others is an essential building block of knowledge production (Grafton, 1999) and serves several purposes (e.g., Neville, 2007, Chapter 2). For example, it contributes to the scientific integrity of the authors by acknowledging other's work and avoiding the impression that they are presenting their own work. Similarly, referencing contributes to the credibility and authenticity of authors' claims. Therefore, it is important to maintain the integrity of what a source says (Ballenger, 2017, Chapter 3), which means that sources should be cited reliably, respectfully, and responsibly (Harvey, 2008; Turabian, 2013). However, referencing in Dash and Paul's (2021) study does not always adhere to these principles as we demonstrate below.

On a number of occasions, Dash and Paul (2021) cite the work of others unreliably, i.e. for something that is not stated in the source. For example, Dash and Paul (2021, p. 1) write that "[r]ecently, the PLS approach has become quite popular among researchers due to its variance based relationship rather than covariance". To support their statement, they cite Mueller and Hancock (2018) and Hayes et al. (2017), among others. However, neither Mueller and Hancock (2018) nor Hayes et al. (2017) refer to PLS(-SEM) in these articles, thus the citations do not support Dash and Paul's claim. Similarly, Dash and Paul (2021, p. 1) mention that "[PLS-SEM] usually deals with a large sample" referring to Mueller and Hancock (2018), Hair et al. (2017a), Hayes et al. (2017), Ullman and Bentler (2003). However, none of these sources refer to PLS-SEM. Consequently, these articles give no evidence that PLS-SEM is commonly used for large samples. In fact, citing (Hair et al., 2013, p. 2, emphasis added), "PLS-SEM has *an erroneous reputation* for offering special sampling capabilities that no other multivariate analysis tool has". Although PLS-SEM will even produce parameter estimates in cases of very small sample sizes, the value and utility of such estimates beyond simple data description, should be questioned (Rigdon, 2016). As a last example of unreliable referencing, Dash and Paul (2021, p. 3) state that "[t]o check the normality of the data, skewness and kurtosis are assessed. The ideal range for these two measures lies between −2 to +2". To support the suggested range for these two measures, they refer to Hu and Bentler (1999). However, Hu and Bentler (1999) do not provide any recommendations for the values of skewness and kurtosis to assess normality. What Hu and Bentler (1999) examine in their article, is cutoff values of various goodness of fit measures under different conditions, including conditions involving non-normally distributed common factors and random measurement errors. For more examples of misplaced references, please see the Appendix.

On other occasions, Dash and Paul (2021) make statements without mentioning a source. Such statements are difficult for a reader to follow or could wrongly create the impression that the idea captured in the statement can genuinely be attributed to Dash and Paul (2021). For example, they state "[the SRMR] is calculated by dividing the fitted residuals by the standard error of the residual" (Dash and Paul, 2021, p. 3). According to Hu and Bentler (1998), and in contrast to Dash and Paul (2021), the SRMR is calculated based on the standardized residuals, which are the differences between the sample and model-implied covariances divided by the standard deviations of the respective observed variables. Thus, to better understand how Dash and Paul calculate the SRMR, a source supporting their claim would be helpful. Similarly, the authors mention "[p]revious studies had always emphasized the limitations posed by regression methods" (Dash and Paul, 2021, p. 8). Yet, these previous studies are not given; therefore, it is difficult to understand the limitations of regression methods. Furthermore, statements such as "[the RMSEA] is considered the best informative fit index" (p. 3) and "[the NNFI] is not affected by the low sample size" (p. 3) would benefit from a reference. To our knowledge, there is no "best" fit index, as all indexes serve different purposes (see, e.g., Schermelleh-Engel et al., 2003), and as Bollen (1986) showed, the NNFI is affected by the sample size. In addition, Dash and Paul (2021) state that "CMIN/df value of 3 (in some cases, even up to 5) or less is considered a good model fit measure" (p. 2) and "[a] threshold value [for the NFI] of 0.90 and above suggests a good model fit" (p. 3). Without a reference, it is unclear to the reader whether these recommendations are taken from the literature (and thus hopefully based on some kind of justification) or whether they are being proposed by Dash and Paul.

*Contributing factor 2: Imprecise and ambiguous language*

Clarity, precision, and accuracy are important standards in the writing of scientific articles (van Thiel, 2021; Kalin et al., 2010; Bown and Gawthorpe, 2010; Blocken, 2017). These standards include avoiding vague, ambiguous, and imprecise language. In contrast, Dash and Paul (2021) use vague and ambiguous language on numerous occasions, which makes their article difficult to understand or leads to confusion. In the following paragraphs, we refer to four examples that illustrate such violation of the principles of clarity, precision, and accuracy in writing scientific articles.

First, according to Dash and Paul's (2021, p. 2) first research question, the purpose of their article is to examine "[w]hich one is the better approach: CB-based model or a PLS-based model". It is inaccurate to equate an approach with a model. An approach, e.g., an estimator, can be applied to different types of models. Moreover, there is not just one type of model for which CB-SEM can be used. In fact, different models can be estimated by CB-SEM because it can emulate various methods that are based on the general linear model such as analysis of variance, regression analysis, and path analysis (Bollen, 1989; Kline, 2015; Bagozzi, 1977). The same applies for PLS-SEM (e.g., Tenenhaus et al., 2005). Although it has been mathematically proven that PLS-SEM is a consistent estimator for structural equation models containing emergent variables (Dijkstra, 2017), in the past, PLS-SEM has (erroneously) been promoted for estimating structural equation models with latent variables in the form of reflective and formative measurement models (e.g., Sarstedt et al., 2014; Hair et al., 2014). Nowadays, it is well-known that PLS-SEM is not suitable for estimating the parameters of latent variable models, i.e., models that comprise reflective and causal-formative measurement models (e.g., Henseler, 2017). Although CB-SEM is predominantly used to estimate latent variable models (see various textbooks such as Bollen, 1989; Kline, 2015), it is also possible to estimate structural equation models containing emergent variables using CB-SEM (e.g., Hancock et al., 2013; Grace and Bollen, 2008; Henseler, 2020; Schuberth, 2021b). With this in mind, it is not clear what Dash and Paul mean when they refer to CB-based and PLS-based models.

Second, Dash and Paul (2021, p. 1) state "CB-SEM models are better for factor-based models like ours, whereas composite-based models provide excellent outcomes in PLS-SEM" . In addition to the confusion surrounding CB-SEM and PLS-SEM models, it is unclear why a particular *model*, e.g., a CB-SEM model (whatever that is), is better for another *model*, e.g., a factor-based model. Models are a means to an end, not an end to a means. Also, it is unclear what the terms 'better' and 'excellent outcome' mean in this context.

Third, the authors say "[the PLS approach] is different from the traditional multivariate techniques that address only individual objectives; It also verifies alternate models to find the most appropriate relationship among the latent variables" (Dash and Paul, 2021, p. 1). Not only is it unclear what they mean by addressing only individual objectives, but it is also not clear why or how PLS verifies alternate models to find the most appropriate relationship between latent variables. To the best of our knowledge, information-theoretic model selection criteria such as Akaike's (1974) information criterion (AIC), which can be used to compare competing models, are far more established in the CB-SEM literature than in the PLS-SEM literature (Sharma et al., 2019; West et al., 2012). Furthermore, as explained in the previous section, PLS-SEM is not suitable for estimating models with latent

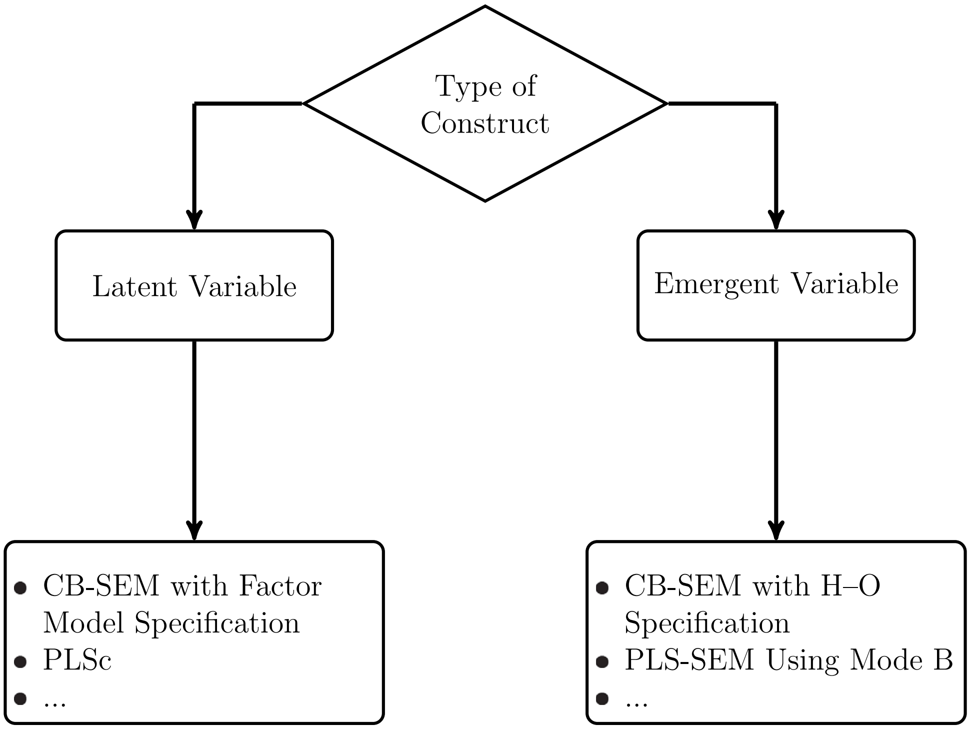



**Fig. 5.** Decision tree for choosing suitable SEM approaches.

variables. Consequently, it is questionable whether PLS-SEM is useful for finding "the most appropriate relationship among latent variables". Similarly, Dash and Paul leave the reader in the dark as to what the term "appropriate relationship" means.

Fourth, there is the statement that "[i]f the numbers vary, e.g., one item with five levels and another with seven levels, it becomes difficult to calculate [the RMSR]" (Dash and Paul, 2021, p. 3). It is unclear why the number of levels of a variable would complicate the calculation of the root mean squared residual (RMSR). The RMSR is defined as the square root of the mean of the squared residuals, where the residuals are the elements of the difference between the sample and the estimated model-implied variance–covariance matrix of the observed variables (e.g., Schermelleh-Engel et al., 2003). Consequently, the number of levels of an observed variable does not affect the calculation of the RMSR.

## 4. Discussion and conclusions

Recently, an article authored by Dash and Paul (2021) appeared in a special issue co-edited by Justin Paul, the second author of the particular article. The article's central conclusion is that "CB or PLS or PLSc do not matter" (p. 9). As our commentary showed both conceptually and by means of a small simulation study, their conclusion is wrong. The choice of SEM approach has a substantial impact on the results of an empirical study. We can, therefore, conclude that it does matter whether CB-SEM, PLS-SEM or PLSc is employed.

In general, analysts should choose an SEM approach that fits their conceptual model. In particular, they should ensure that the chosen approach to SEM allows them to estimate the models containing their specified constructs (see Fig. 5). To estimate the parameters associated with a latent variable, analysts can use CB-SEM with a factor model specification or PLSc. In contrast, for parameters associated with an emergent variable, analysts can use CB-SEM in combination with the H–O specification or PLS-SEM using Mode B. If a structural equation model contains both latent and emergent variables, CB-SEM with a mixture of the H–O and factor model specifications, or PLSc, where no correction for attenuation is applied for emergent variables, are viable options.

## CRediT authorship contribution statement

**Florian Schuberth:** Investigation, Supervision, Writing — original draft. **Geoffrey Hubona:** Writing – original draft. **Ellen Roemer:** Writing – original draft. **Sam Zaza:** Writing – original draft. **Tamara Schamberger:** Methodology, Formal analysis, Writing – review & editing. **Francis Chuah:** Writing – review & editing. **Gabriel Cepeda-Carrión:** Writing – review & editing. **Jörg Henseler:** Conceptualization, Funding acquisition, Writing – original draft.

## Data availability

The two population datasets and the R code to conduct the analysis can be downloaded from the following URL: https://osf.io/z2h7v/?view_only=52e0d1e574dc42e2a22b92192d87c16a.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.techfore.2023.122665.

## References

Aguirre-Urreta, M.I., Marakas, G.M., 2014a. A rejoinder to Rigdon et al. (2014). Inf. Syst. Res. 25 (4), 785–788. http://dx.doi.org/10.1287/isre.2014.0545.
Aguirre-Urreta, M.I., Marakas, G.M., 2014b. Research note: Partial least squares and models with formatively specified endogenous constructs: A cautionary note. Inf. Syst. Res. 25 (4), 761–778. http://dx.doi.org/10.1287/isre.2013.0493.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19 (6), 716–723. http://dx.doi.org/10.1109/tac.1974.1100705.
Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge Massachusetts.
Bagozzi, R.P., 1977. Structural equation models in experimental research. J. Mar. Res. 14 (2), 209–226. http://dx.doi.org/10.1177/002224377701400209.
Ballenger, B.P., 2017. The Curious Researcher: A Guide to Writing Research Papers, ninth ed. Pearson, Hoboken, NJ.
Benitez, J., Henseler, J., Castillo, A., Schuberth, F., 2020. How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. Inf. Manage. 2 (57), 103168. http://dx.doi.org/10.1016/j.im.2019.05.003.

Blaikie, N., 2000. Designing Social Research: The Logic of Anticipation. Polity, Malden.
Blocken, B., 2017. 10 Tips for writing a truly terrible journal article. Elsevier Connect Website URL: https://www.elsevier.com/connect/authors-update/10-tips-for-writing-a-truly-terrible-journal-article.
Bollen, K.A., 1986. Sample size and Bentler and Bonett's nonnormed fit index. Psychometrika 51 (3), 375–377. http://dx.doi.org/10.1007/BF02294061.
Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.
Bollen, K.A., 2002. Latent variables in psychology and the social sciences. Annu. Rev. Psychol. 53 (1), 605–634. http://dx.doi.org/10.1146/annurev.psych.53.100901.135239.
Bollen, K.A., Bauldry, S., 2011. Three Cs in measurement models: Causal indicators, composite indicators, and covariates. Psychol. Methods 16 (3), 265–284. http://dx.doi.org/10.1037/a0024448.
Bollen, K.A., Diamantopoulos, A., 2017. In defense of causal–formative indicators: A minority report. Psychol. Methods 22 (3), 581–596. http://dx.doi.org/10.1037/met0000056.
Bono, J.E., McNamara, G., 2011. Publishing in AMJ–part 2: Research design. Acad. Manag. J. 54 (4), 657–660. http://dx.doi.org/10.5465/amj.2011.64869103.
Bown, L., Gawthorpe, A., 2010. Get Your Articles Published, second ed. Teach Yourself.
Broad, W.J., Wade, N., 1982. Betrayers of the Truth. Simon & Schuster, New York.
Browne, M.W., 1984. Asymptotically distribution-free methods for the analysis of covariance structures. Br. J. Math. Stat. Psychol. 37 (1), 62–83. http://dx.doi.org/10.1111/j.2044-8317.1984.tb00789.x.
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, 243–272. http://dx.doi.org/10.1007/s41237-019-00098-0.
Cho, G., Hwang, H., Sarstedt, M., 2022. A comparative evaluation of factor- and component-based structural equation modeling approaches under (in)correct construct representations. Br. J. Math. Stat. Psychol. 75 (2), 220–251. http://dx.doi.org/10.1111/bmsp.12255.
Dash, G., Paul, J., 2021. CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. Technol. Forecast. Soc. Change 173, 121092. http://dx.doi.org/10.1016/j.techfore.2021.121092.
de Rooij, M., Karch, J.D., Fokkema, M., Bakk, Z., Pratiwi, B.C., Kelderman, H., 2022. SEM-based out-of-sample predictions. Struct. Equ. Model. 1–17. http://dx.doi.org/10.1080/10705511.2022.2061494.
Dijkstra, T.K., 1981. Latent variables in linear stochastic models: Reflections on "Maximum Likelihood" and "Partial Least Squares" methods (Ph.D. thesis). Groningen University.
Dijkstra, T.K., 2017. A perfect match between a model and a mode. In: Latan, H., Noonan, R. (Eds.), Partial Least Squares Path Modeling. Springer, Cham, pp. 55–80. http://dx.doi.org/10.1007/978-3-319-64069-3_4.
Dijkstra, T.K., Henseler, J., 2015a. Consistent and asymptotically normal PLS estimators for linear structural equations. Comput. Statist. Data Anal. 81, 10–23. http://dx.doi.org/10.1016/j.csda.2014.07.008.
Dijkstra, T.K., Henseler, J., 2015b. Consistent partial least squares path modeling. MIS Q. 39 (2), 297–316. http://dx.doi.org/10.25300/MISQ/2015/39.2.02.
Dijkstra, T.K., Schermelleh-Engel, K., 2014. Consistent partial least squares for nonlinear structural equation models. Psychometrika 79 (4), 585–604. http://dx.doi.org/10.1007/S11336-013-9370-0.
Evermann, J., Rönkkö, M., 2021. Recent developments in PLS. Commun. Assoc. Inf. Syst. in press.
Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Mar. Res. 18 (1), 39–50. http://dx.doi.org/10.1177/002224378101800104.
Gefen, D., Straub, D.W., Rigdon, E.E., 2011. An update and extension to SEM guidelines for administrative and social science research. MIS Q. 35 (2), iii–xiv. http://dx.doi.org/10.2307/23044042.
Goodhue, D.L., Lewis, W., Thompson, R., 2012. Comparing PLS to regression and LISREL: A response to Marcoulides, Chin, and Saunders. MIS Q. 36 (3), 703–716. http://dx.doi.org/10.2307/41703476.
Goodhue, D.L., Lewis, W., Thompson, R., 2017. A multicollinearity and measurement error statistical blind spot. MIS Q. 41 (3), 667–A15. http://dx.doi.org/10.25300/MISQ/2017/41.3.01.
Goodstein, D., 2010. On Fact and Fraud. Princeton University Press, Princeton, NJ.
Grace, J.B., Bollen, K.A., 2008. Representing general theoretical concepts in structural equation models: The role of composite variables. Environ. Ecol. Stat. 15 (2), 191–213. http://dx.doi.org/10.1007/s10651-007-0047-7.
Grafton, A., 1999. The Footnote: A Curious History. Harvard University Press, Cambridge, MA.
Hair, J.F., Babin, B.J., Krey, N., 2017a. Covariance-based structural equation modeling in the Journal of Advertising: Review and recommendations. J. Advert. 46 (3), 454. http://dx.doi.org/10.1080/00913367.2017.1329496.
Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Thiele, K.O., 2017b. Mirror, mirror on the wall: a comparative evaluation of composite-based structural equation modeling methods. J. Acad. Mark. Sci. 45 (5), 616–632. http://dx.doi.org/10.1007/s11747-017-0517-x.
Hair, J.F., Ringle, C.M., Sarstedt, M., 2013. Editorial: Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long Range Plan. 46 (1), 1–12. http://dx.doi.org/10.1016/j.lrp.2013.01.001.
Hair, J.F., Sarstedt, M., Hopkins, L., Kuppelwieser, V.G., 2014. Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. Eur. Bus. Rev. 26 (2), 106–121. http://dx.doi.org/10.1108/ebr-10-2013-0128.
Hancock, G.R., Mao, X., Kher, H., 2013. On latent growth models for composites and their constituents. Multivar. Behav. Res. 48 (5), 619–638. http://dx.doi.org/10.1080/00273171.2013.815579.
Harvey, G., 2008. Writing with Sources: A Guide for Students, second ed. Hackett, Indianapolis.
Hayes, A.F., Montoya, A.K., Rockwood, N.J., 2017. The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. Australas. Mark. J. 25 (1), 76–81. http://dx.doi.org/10.1016/j.ausmj.2017.02.001.
Henseler, J., 2017. Bridging design and behavioral research with variance-based structural equation modeling. J. Advert. 46 (1), 178–192. http://dx.doi.org/10.1080/00913367.2017.1281780.
Henseler, J., 2018. Partial least squares path modeling: Quo vadis? Qual. Quant. 52 (1), 1–8. http://dx.doi.org/10.1007/s11135-018-0689-6.
Henseler, J., 2020. Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables. Guilford Press, New York, NY.
Henseler, J., Schuberth, F., 2020a. Auxiliary theories. In: Henseler, J. (Ed.), Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables. The Guilford Press, London, New York, pp. 25–37.
Henseler, J., Schuberth, F., 2020b. Using confirmatory composite analysis to assess emergent variables in business research. J. Bus. Res. 120, 147–156. http://dx.doi.org/10.1016/j.jbusres.2020.07.026.
Hu, L.t., Bentler, P.M., 1998. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychol. Methods 3 (4), 424–453. http://dx.doi.org/10.1037/1082-989X.3.4.424.
Hu, L.t., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equ. Model. 6 (1), 1–55. http://dx.doi.org/10.1080/10705519909540118.
Hui, B.S., Wold, H., 1982. Consistency and consistency at large of partial least squares estimates. In: Wold, H.O.A., Jöreskog, K.G. (Eds.), Systems under Indirect Observation: Causality, Structure, Prediction, Vol. II. North-Holland, Amsterdam, pp. 119–130.
Hwang, H., Cho, G., Jung, K., Falk, C., Flake, J., Jin, M.J., Lee, S.H., 2021. An approach to structural equation modeling with both factors and components: Integrated generalized structured component analysis. Psychol. Methods 26 (3), 273–294. http://dx.doi.org/10.1037/met0000336.
Jöreskog, K.G., 1970. A general method for estimating a linear structural equation system. ETS Res. Bull. Ser. 1970 (2), i–41. http://dx.doi.org/10.1002/j.2333-8504.1970.tb00783.x.
Jöreskog, K.G., 1971b. Statistical analysis of sets of congeneric tests. Psychometrika 36 (2), 109–133. http://dx.doi.org/10.1007/bf02291393.
Kalin, M., Weygant, R.S., Rosen, H.J., Regener, J.R., 2010. Construction Specifications Writing: Principles and Procedures, sixth ed. Wiley, Hoboken, NJ.
Kline, R.B., 2015. Principles and Practice of Structural Equation Modeling, fourth ed. The Guilford Press, New York.
Lohmöller, J.B., 1989. Latent Variable Path Modeling with Partial Least Squares. Physica, Heidelberg.
McIntosh, C.N., Edwards, J.R., Antonakis, J., 2014. Reflections on partial least squares path modeling. Organ. Res. Methods 17 (2), 210–251. http://dx.doi.org/10.1177/1094428114529165.
Meibauer, J., 2016. Aspects of a theory of bullshit. Pragmat. Cogn. 23 (1), 68—91. http://dx.doi.org/10.1075/pc.23.1.04mei.
Mueller, R.O., Hancock, G.R., 2018. Structural equation modeling. In: Hancock, G.R., Stapleton, L.M., Mueller, R.O. (Eds.), The Reviewer's Guide to Quantitative Methods in the Social Sciences, second ed. Routledge, New York, pp. 445–456.
Neville, C., 2007. The Complete Guide to Referencing and Avoiding Plagiarism. Open University Press, London.
Paxton, P., Curran, P.J., Bollen, K.A., Kirby, J., Chen, F., 2001. Monte Carlo experiments: Design and implementation. Struct. Equ. Model. 8 (2), 287–312. http://dx.doi.org/10.1207/S15328007SEM0802_7.
R. Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.
Rademaker, M.E., Schuberth, F., 2020. cSEM: Composite-based structural equation modeling. URL: https://m-e-rademaker.github.io/cSEM/, Package version: 0.5.0.9000.
Rademaker, M.E., Schuberth, F., Dijkstra, T.K., 2019. Measurement error correlation within blocks of indicators in consistent partial least squares: Issues and remedies. Internet Res. 29 (3), 448–463. http://dx.doi.org/10.1108/IntR-12-2017-0525.
Reinartz, W., Haenlein, M., Henseler, J., 2009. An empirical comparison of the efficacy of covariance-based and variance-based SEM. Int. J. Res. Mark. 26 (4), 332–344. http://dx.doi.org/10.1016/j.ijresmar.2009.08.001.
Rigdon, E.E., 2016. Choosing PLS path modeling as analytical method in European management research: A realist perspective. Eur. Manag. J. 34 (6), 598–605. http://dx.doi.org/10.1016/j.emj.2016.05.006.
Rigdon, E.E., Becker, J.M., Rai, A., Ringle, C.M., Diamantopoulos, A., Karahanna, E., Straub, D.W., Dijkstra, T.K., 2014. Conflating antecedents and formative indicators: A comment on Aguirre-Urreta and Marakas. Inf. Syst. Res. 25 (4), 780–784. http://dx.doi.org/10.1287/isre.2014.0543.

Rönkkö, M., Evermann, J., 2013. A critical examination of common beliefs about partial least squares path modeling. Organ. Res. Methods 16 (13), 425–448. http://dx.doi.org/10.1177/1094428112474693.

Rönkkö, M., McIntosh, C.N., Antonakis, J., Edwards, J.R., 2016. Partial least squares path modeling: Time for some serious second thoughts. J. Oper. Manage. 47–48, 9–27. http://dx.doi.org/10.1016/j.jom.2016.05.002.

Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. J. Stat. Softw. 48 (2), 1–36. http://dx.doi.org/10.18637/jss.v048.i02.

Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O., Gudergan, S.P., 2016. Estimation issue with PLS and CBSEM: Where the bias lies! J. Bus. Res. 69 (10), 3998–4010. http://dx.doi.org/10.1016/j.jbusres.2016.06.007.

Sarstedt, M., Ringle, C.M., Smith, D., Reams, R., Hair, J.F., 2014. Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. J. Fam. Bus. Strategy 5 (1), 105–115. http://dx.doi.org/10.1016/j.jfbs.2014.01.002.

Schamberger, T., 2023. Conducting Monte Carlo simulations with PLS-PM and other variance-based estimators for structural equation models. Ind. Manage. Data Syst. 123 (6), 1789–1813. http://dx.doi.org/10.1108/IMDS-07-2022-0418.

Schermelleh-Engel, K., Moosbrugger, H., Müller, H., 2003. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods Psychol. Res. Online 8 (2), 23–74.

Schuberth, F., 2021a. Confirmatory composite analysis using partial least squares: Setting the record straight. Rev. Manag. Sci. 15, 1311–1345. http://dx.doi.org/10.1007/s11846-020-00405-0.

Schuberth, F., 2021b. The Henseler–Ogasawara specification of composites in structural equation modeling: A tutorial. Psychol. Methods http://dx.doi.org/10.1037/met0000432.

Schuberth, F., Henseler, J., Dijkstra, T.K., 2018b. Partial least squares path modeling using ordinal categorical indicators. Qual. Quant. 52 (1), 9–35. http://dx.doi.org/10.1007/s11135-016-0401-7.

Schuberth, F., Zaza, S., Henseler, J., 2023. Partial least squares is an estimator for structural equation models: A comment on Evermann and Rönkkö (2021). Commun. Assoc. Inf. Syst. (in press).

Sharma, P., Sarstedt, M., Shmueli, G., Kim, K., Thiele, K., 2019. PLS-based model selection: The role of alternative explanations in information systems research. J. Assoc. Inf. Syst. 20 (4), 346–397. http://dx.doi.org/10.17005/1.jais.00538.

Shmueli, G., Ray, S., Estrada, J.M.V., Chatla, S.B., 2016. The elephant in the room: Predictive performance of PLS models. J. Bus. Res. 69 (10), 4552–4564. http://dx.doi.org/10.1016/j.jbusres.2016.03.049.

Stroebe, W., Postmes, T., Spears, R., 2012. Scientific misconduct and the myth of self-correction in science. Perspect. Psychol. Sci. 7 (6), 670–688. http://dx.doi.org/10.1177/1745691612460687.

Tenenhaus, M., Vinzi, V.E., Chatelin, Y.M., Lauro, C., 2005. PLS path modeling. Comput. Statist. Data Anal. 48 (1), 159–205. http://dx.doi.org/10.1016/j.csda.2004.03.005.

Turabian, K.L., 2013. A Manual for Writers of Research Papers, Theses, and Dissertations: Chicago Style for Students and Researchers, eighth ed. University of Chicago Press, Chicago.

Ullman, J.B., Bentler, P.M., 2003. Structural equation modeling. In: Schinka, J.A., Velicer, W.F. (Eds.), Handbook of Psychology. Wiley, Hoboken, NJ, pp. 607–634.

van Thiel, S., 2021. Research Methods in Public Administration and Public Management an Introduction, second ed. Routledge, London.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth Springer, New York.

Werts, C.E., Linn, R.L., Jöreskog, K.G., 1974. Intraclass reliability estimates: Testing structural assumptions. Educ. Psychol. Meas. 34 (1), 25–33. http://dx.doi.org/10.1177/001316447403400104.

West, S.G., Taylor, A.B., Wu, W., 2012. Model fit and model selection in structural equation modeling. In: Hoyle, R.H. (Ed.), Handbook of Structural Equation Modeling. Guilford Press, New York, pp. 209–231.

Wold, H.O.A., 1975. Path models with latent variables: The NIPALS approach. In: Blalock, H.M., Aganbegian, A., Borodkin, F., Boudon, R., Capecchi, V. (Eds.), Quantitative Sociology. Academic Press, Boston, MA, pp. 307–357.

Wooldridge, J.M., 2012. Introductory Econometrics: A Modern Approach, fifth ed. Cengage Learning, Boston, MA.

Yu, X., Schuberth, F., Henseler, J., 2023. Specifying composites in structural equation modeling: A refinement of the Henseler–Ogasawara specification. Stat. Anal. Data Min. http://dx.doi.org/10.1002/sam.11608.

Yu, X., Zaza, S., Schuberth, F., Henseler, J., 2021. Counterpoint: Representing forged concepts as emergent variables using composite-based structural equation modeling. ACM SIGMIS Database: DATABASE Adv. Inf. Syst. 52 (SI), 114–130. http://dx.doi.org/10.1145/3505639.3505647.

## Further reading

Beran, R., Srivastava, M.S., 1985. Bootstrap tests and confidence regions for functions of a covariance matrix. Ann. Statist. 13 (1), 95–115. http://dx.doi.org/10.1214/aos/1176346579.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., Billiet, J., 2014. Measurement equivalence in cross-national research. Annu. Rev. Sociol. 40, 55–75. http://dx.doi.org/10.1146/annurev-soc-071913-043137.

Gerbing, D.W., Anderson, J.C., 1988. An updated paradigm for scale development incorporating unidimensionality and its assessment. J. Mar. Res. 25 (2), 186–192. http://dx.doi.org/10.1177/002224378802500207.

Jöreskog, K.G., 1971a. Simultaneous factor analysis in several populations. Psychometrika 36 (4), 409–426. http://dx.doi.org/10.1007/bf02291366.

Klesel, M., Schuberth, F., Niehaves, B., Henseler, J., 2022. Multigroup analysis in information systems research using PLS-PM: A systematic investigation of approaches. DATA BASE Adv. Inf. Syst. 53 (3), 26–48. http://dx.doi.org/10.1145/3551783.3551787.

Marsh, H.W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., Trautwein, U., 2009. Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Struct. Equ. Model. 16 (3), 439–476. http://dx.doi.org/10.1080/10705510903008220.

Sarstedt, M., Henseler, J., Ringle, C.M., 2011. Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. In: Sarstedt, M., Schwaiger, M., Taylor, C. (Eds.), Measurement and Research Methods in International Marketing. In: Advances in International Marketing, vol. 22, Emerald Group Publishing Limited, Bingley, pp. 195–218. http://dx.doi.org/10.1108/S1474-7979(2011)0000022012.

Schuberth, F., Henseler, J., Dijkstra, T.K., 2018a. Confirmatory composite analysis. Front. Psychol. 9, 2541. http://dx.doi.org/10.3389/fpsyg.2018.02541.

Schuberth, F., Rademaker, M.E., Henseler, J., 2022. Assessing the overall fit of composite models estimated by partial least squares path modeling. Eur. J. Mark. 57 (6), 1678–1702. http://dx.doi.org/10.1108/EJM-08-2020-0586.

**Florian Schuberth** is Assistant Professor at the Chair of Product–Market Relations in the Faculty of Engineering Technology at the University of Twente, The Netherlands. He obtained his Ph.D. degree in Econometrics at the Faculty of Business Management and Economics, University of Würzburg, Germany. His main research interests are structural equation modeling, in particular on composite-based estimators and their enhancement. He is also co-inventor of confirmatory composite analysis (CCA) and co-developer of cSEM, an R package for composite-based structural equation modeling. His research has been published in various international peer-reviewed journals, including European Journal of Marketing, Information & Management, International Journal of Information Management, Psychological Methods, and Structural Equation Modeling—A Multidisciplinary Journal.

**Geoffrey Hubona** is an associate professor in the A.R. Sanchez School of Business at Texas A&M International University where he teaches various data science and programming courses at the undergraduate and graduate levels. He is an active researcher largely engaged in methodology issues, the user acceptance of information technologies, and cross-cultural studies of technology adoption. He has been published in MIS Quarterly, ACM Transactions on Computer–Human Interaction, IEEE Transactions on Systems, Man and Cybernetics, Data Base for Advances in Information Systems, Industrial Data & Management Systems, Journal of Services Marketing, International Journal of Information Management, Information & Management, Information Technology & People, Journal of Global Information Management, International Journal of Human Computer Studies, International Journal of Technology and Human Interaction, Journal of Organizational and End User Computing, Journal of Information Technology Management, Kybernetes, and Data Analysis Perspectives Journal.

**Ellen Roemer** is Professor of Market Research and International Marketing at Hochschule Ruhr West, University of Applied Sciences, Mülheim an der Ruhr, Germany. Her research interests focus on the analysis of customer relationships, with special attention toward the adoption and acceptance of eco-innovations. She complements her theoretical work with empirical studies using qualitative and quantitative research designs. Noteworthily, she uses partial least squares path modeling or experimental designs with eye-tracking to answer research questions. She has published in leading journals such as Industrial Data and Management Systems, Industrial Marketing Management, Journal of Marketing Management and Journal of Strategic Marketing.

**Sam Zaza** is an Assistant Professor at the Information Systems and Analytics Department, Middle Tennessee State University. She earned her Ph.D. in Management Information Systems from Florida State University. Sam's research interests span Human–Computer Interactions, IT Workforce, Diversity and Inclusion challenges and initiatives, IT artifact, and methodological approaches. Sam has published (or forthcoming) articles in Management Information Systems Quarterly, Information and Organization, IT & People, and Communications of AIS, among others. Sam participates and presents in major IS conferences, internationally and regionally. Sam is currently serving as the AIS Women's Network Chair, SIG Social Inclusion President, SIG Lead President, and Past President of the Middle East and North Africa-AIS.

**Tamara Schamberger** is a Postdoctoral Researcher at the Faculty of Business Management and Economics at the University of Würzburg, Germany. She obtained her Ph.D. degree in Econometrics in a joint doctorate at the Faculty of Business Management and Economics, University of Würzburg, Germany and the Department of Design, Production and Management of the University of Twente, the Netherlands. Her main research interests are on composite-based structural equation modeling, in particular on methodological enhancements. Her research has been published in various international peer-reviewed journals such as Behaviormetrika, International Journal of Behavioral Development or the British Journal of Mathematical and Statistical Psychology.

**Francis Chuah** is a senior lecturer in the School of Business Management, Universiti Utara Malaysia. He holds a Ph.D. in Management from Putra Business School, Universiti Putra Malaysia and is currently co-managing editor of the Journal of Applied Structural Equation Modelling (JASEM). He serves as an expert panel for Sarawak Research Society (SRS), a non-profit making and scholarly body in Sarawak and has been a professional facilitator for Tunku Abdul Rahman University College, Universiti Teknologi Mara (UiTM) Sabah and Multimedia University (MMU). He serves as a reviewer for several high impact journals indexed by Scopus and Web of Science in the field of business and management. His research interest is in strategy and management, with main emphasis on behavioral strategy. He is also interested in research related to methodological advancement. He teaches strategic management, research methodology and quantitative data analysis at both undergraduate and postgraduate levels. He is well-versed in the application of partial least squares structural equation modeling technique and has conducted workshops at both basic and advanced levels.

**Gabriel Cepeda-Carrión** is a Professor of Management in the Department of Management and Marketing, at Universidad de Sevilla, Spain. His research interests include dynamic capabilities, knowledge management, leadership, sport management, and partial least squares structural equation modeling (PLS-SEM). His recent contributions have been published in Journal of Knowledge Management, Internet Research, Industrial Management & Data Systems, European Management Journal, Management Decision, British Journal of Management, Journal of Business Research, Journal of Organizational Change Management, Journal of Services Marketing, and Total Quality Management & Business Excellence, among others. He is currently on the editorial board of the International Journal of Sports Marketing and Sponsorship. He has also served as Guest Editor of the Journal of Business Research, European Management Journal, and International Journal of Sports Marketing and Sponsorship. In addition, Dr. Cepeda has been conference co-chair of the 2nd International Symposium on Partial Least Squares Path Modeling—The Conference for PLS Users (June 16–19, Seville, Spain) and the 9th International Conference on PLS and Related Methods (PLS'201917) (June 17–19, 2017, Macau, China).

**Jörg Henseler** is Professor of Product–Market Relations at the University of Twente, the Netherlands, and a visiting professor at NOVA Information Management School, Universidade Nova de Lisboa, Portugal. He is co-inventor of consistent partial least squares (PLSc), the heterotrait–monotrait ratio of correlations (HTMT), the H–O specification, and confirmatory composite analysis (CCA). He authored the monograph "Composite-based Structural Equation Modeling: Analyzing Latent and Emergent Variables" (Guilford Press, 2020). According to Clarivate, he is a highly cited researcher (2018–), and in 2022 he was ranked among the 500 most impactful researchers in the world by J. Ioannides (Stanford). His more than 100 scholarly articles have been published in journals such as Computational Statistics & Data Analysis, Journal of the Academy of Marketing Science, MIS Quarterly, Organizational Research Methods, and Structural Equation Modeling—A Multidisciplinary Journal, among others. He chairs the Scientific Advisory Board of ADANCO, software for composite-based structural equation modeling (https://www.composite-modeling.com).