Assessing the overall fit of composite models estimated by partial least squares path modeling

Florian Schuberth
*Faculty of Engineering Technology, Universiteit Twente,
Enschede, The Netherlands*

Manuel E. Rademaker
*Faculty of Business Management and Economics,
Julius-Maximilians-Universität Würzburg, Würzburg, Germany, and*

Jörg Henseler
*Faculty of Engineering Technology, Universiteit Twente, Enschede,
The Netherlands and Nova Information Management School,
Universidade Nova de Lisboa, Lisbon, Portugal*

Abstract

Purpose – This study aims to examine the role of an overall model fit assessment in the context of partial least squares path modeling (PLS-PM). In doing so, it will explain when it is important to assess the overall model fit and provides ways of assessing the fit of composite models. Moreover, it will resolve major concerns about model fit assessment that have been raised in the literature on PLS-PM.

Design/methodology/approach – This paper explains when and how to assess the fit of PLS path models. Furthermore, it discusses the concerns raised in the PLS-PM literature about the overall model fit assessment and provides concise guidelines on assessing the overall fit of composite models.

Findings – This study explains that the model fit assessment is as important for composite models as it is for common factor models. To assess the overall fit of composite models, researchers can use a statistical test and several fit indices known through structural equation modeling (SEM) with latent variables.

Research limitations/implications – Researchers who use PLS-PM to assess composite models that aim to understand the mechanism of an underlying population and draw statistical inferences should take the concept of the overall model fit seriously.

Practical implications – To facilitate the overall fit assessment of composite models, this study presents a two-step procedure adopted from the literature on SEM with latent variables.

Originality/value – This paper clarifies that the necessity to assess model fit is not a question of which estimator will be used (PLS-PM, maximum likelihood, etc). but of the purpose of statistical modeling. Whereas, the model fit assessment is paramount in explanatory modeling, it is not imperative in predictive modeling.

Keywords Partial least squares path modeling, Composite models, Emergent variables, Goodness of fit, Model fit assessment

Paper type Technical paper

1. Introduction

Over the past decade, composite models have drawn increasing interest in the context of structural equation modeling (SEM). The composite model is regarded as a viable alternative to the common factor model as a means to operationalize and relate abstract concepts from marketing and other disciplines (Sarstedt *et al.*, 2016; Henseler, 2021). Different from the common factor model, the composite model expresses abstract concepts by emergent variables, i.e. composites of observed variables, instead of latent variables [1]. To interrelate emergent variables, two different models have been suggested. First, a model where emergent variables freely correlate (Schuberth *et al.*, 2018). Second, a model where emergent variables are embedded in a structural model (Dijkstra, 2017).

Arguably, the most widespread estimator for composite models is partial least squares path modeling (PLS-PM; Wold, 1982). Its statistical properties are well studied (Dijkstra, 1985), and its use is appreciated by researchers across various fields, including marketing (Hair *et al.*, 2012). PLS-PM can be used for various types of research (Henseler, 2018), and various enhancements have been developed over the past decade (Khan *et al.*, 2019). Moreover, PLS-PM has been implemented in various statistical software packages, including commercial software such as ADANCO (Henseler and Dijkstra, 2017) or SmartPLS (Ringle *et al.*, 2015) and open-source packages such as cSEM (Rademaker and Schuberth, 2020).

In SEM with latent variables, the overall model fit assessment is considered to be a crucial step (Kline, 2015). This assessment investigates whether the specified model is consistent with the data collected by exploiting constraints imposed on the observed variables' model-implied variance–covariance matrix, in line with the maxim that, "[i]f a model is consistent with reality, then the data should be consistent with the model" (Bollen, 1989, p. 68). The overall model fit can be assessed in two nonexclusive ways, namely, tests for the overall model fit and fit indices (Schermelleh-Engel *et al.*, 2003). While the former are based on statistical inference, the latter are usually descriptive and quantify the misfit on a continuous scale.

Currently, the literature on PLS-PM takes divergent stands on the overall fit assessment. While proponents mainly follow the reasoning known from SEM with latent variables (Henseler *et al.*, 2016; Henseler, 2017; Benitez *et al.*, 2020), critics, dating back to Lohmöller (1989), have raised several concerns about the model fit assessment, which circle around the following six arguments:

- PLS-PM is focused on estimating causal–predictive relationships (Hair *et al.*, 2020).

- Assessing the model fit by means of a distance function is not useful in the context of PLS-PM (Hair *et al.*, 2017, 2019a).

- Fit indices based on the common factor model are not appropriate to assess a model estimated by PLS-PM (Lohmöller, 1989, p. 54).

- Thresholds for fit indices have not been proposed for composite models (Hair *et al.*, 2019b).

- It is unclear whether the fit should be assessed based on the estimated model or on a model with a saturated structural model (Hair *et al.*, 2019b).

- Small misspecifications are not reliably detected by the bootstrap-based test for the overall model fit (Hair *et al.*, 2020).

Apparently, it is not clear *when* the overall fit of composite models needs to be assessed, and if so, *how* it should be assessed, thereby giving rise to confusion.

In light of this situation, our paper contributes to the PLS-PM literature in five ways. First, we clarify in which research situations the overall fit assessment of composite models estimated by

PLS-PM is paramount. Second, we provide an overview of the available means of assessing the overall fit of composite models estimated by PLS-PM, i.e. a bootstrap-based test and various fit indices. Third, we address the concerns about model fit assessment raised in the PLS-PM literature. Fourth, by means of a Monte Carlo simulation, we demonstrate the finite-sample behavior of the bootstrap-based test for the overall model fit in combination with various fit measures and show that it is able to detect misspecified composite models. Fifth, we provide concise guidelines on how to assess the overall fit of composite models. Overall, we answer the questions of when and how to assess the fit of composite models estimated by PLS-PM and show that the raised concerns are mainly unfounded.

2. A formal definition of the composite model

The composite model is a model that is consistently estimated by PLS-PM (Dijkstra, 2017) [2]. It comprises several emergent variables, and each emergent variable η_j is completely determined by a unique block of K_j observed variables, $\eta_j = \boldsymbol{w}_j^\top \boldsymbol{x}_j$, where the vector $\boldsymbol{w}_j$ contains the weights of block j and the vector $\boldsymbol{x}_j$ contains the K_j observed variables of block j. It is assumed that all observed variables are standardized and that each observed variable is connected to only one emergent variable.

In the composite model studied in the context of confirmatory composite analysis (CCA; Schuberth *et al.*, 2018; Henseler and Schuberth, 2020, Hubona *et al.*, 2021), the emergent variables are typically allowed to freely correlate [3]. Hence, the emergent variables' variance–covariance matrix $\boldsymbol{\Phi}$ is unconstrained. The model-implied variance–covariance matrix of the observed variables $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where the vector $\boldsymbol{\theta}$ comprises the model parameters, can be expressed as a partitioned matrix:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1J} \\ & \boldsymbol{\Sigma}_{22} & \dots & \boldsymbol{\Sigma}_{2J} \\ & & \ddots & \vdots \\ & & & \boldsymbol{\Sigma}_{JJ} \end{pmatrix}. \tag{1}$$

The variances and covariances of the observed variables of block j are captured in the *intra-block* variance–covariance matrix $\boldsymbol{\Sigma}_{jj}$. Typically, all observed variables of one block freely correlate. The covariances between the observed variables of blocks j and l are captured in the *inter-block* covariance matrix $\boldsymbol{\Sigma}_{jl}$, with $j \neq l$. In contrast to the intra-block variance–covariance matrices, the inter-block covariance matrices contain constraints, namely, that the emergent variables carry all information between the blocks of observed variables:

$$\boldsymbol{\Sigma}_{jl} = \phi_{jl} \boldsymbol{\Sigma}_{jj} \boldsymbol{w}_j \boldsymbol{w}_l^\top \boldsymbol{\Sigma}_{ll} = \phi_{jl} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_l^\top, \tag{2}$$

where the scalar ϕ_{jl} equals the covariance between the emergent variables η_j and η_l. The covariances between the emergent variable η_j and its respective observed variables $\boldsymbol{x}_j$ are captured in the vector $\boldsymbol{\lambda}_j = \boldsymbol{\Sigma}_{jj} \boldsymbol{w}_j$. The latter are often labeled composite loadings.

Additionally, the emergent variables $\boldsymbol{\eta}$ can be embedded in a structural model (Dijkstra, 2017). Therefore, we distinguish between exogenous emergent variables ($\boldsymbol{\eta}_{exo}$) and endogenous emergent variables ($\boldsymbol{\eta}_{endo}$). The former contains emergent variables that are not explained by other emergent variables in the structural model. Equation (3) provides a formal representation of a linear structural model containing emergent variables:

$$\boldsymbol{\eta}_{endo} = \boldsymbol{\Gamma}\boldsymbol{\eta}_{exo} + \boldsymbol{B}\boldsymbol{\eta}_{endo} + \boldsymbol{\zeta} \tag{3}$$

The matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{B}$ capture the respective coefficients of the exogenous and endogenous emergent variables. The error terms of each structural equation are captured in the vector $\boldsymbol{\zeta}$ and are assumed to have a mean of zero. For simplicity, it is assumed that they are mutually uncorrelated and uncorrelated with the exogenous emergent variables $\boldsymbol{\eta}_{exo}$. Consequently, the variance–covariance matrix of the emergent variables $\boldsymbol{\Phi}$ has the following structure:

$$\boldsymbol{\Phi} = \begin{pmatrix} \text{cov}(\boldsymbol{\eta}_{exo}) & \text{cov}(\boldsymbol{\eta}_{exo})\left((\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\right)^{\top} \\ (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\text{cov}(\boldsymbol{\eta}_{exo})\left((\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\right)^{\top} + (\boldsymbol{I}-\boldsymbol{B})^{-1}\text{cov}(\boldsymbol{\zeta})\left((\boldsymbol{I}-\boldsymbol{B})^{-1}\right)^{\top} \end{pmatrix} \tag{4}$$

The identity matrix $\boldsymbol{I}$ is of the same dimension as the coefficient matrix $\boldsymbol{B}$, and it is assumed that $\boldsymbol{I}-\boldsymbol{B}$ is nonsingular.

The calculation of the model-implied variance–covariance matrix of the observed variables $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where the emergent variables are embedded in a structural model, is similar to that in the case in which the emergent variables are allowed to freely correlate, i.e. the variance–covariance structure has the same form as shown in Equation (1). However, the covariance ϕ_{jl} between the emergent variables $\boldsymbol{\eta}_j$ and $\boldsymbol{\eta}_l$ needs to be replaced by the corresponding element of the variance–covariance matrix of the emergent variables as implied by the structural model shown in Equation (4).

3. When to assess the overall fit of composite models

The type of research question (Leek and Peng, 2015; Henseler, 2018), and thus the purpose of the research, determines the type of statistical modeling. In general, two types of statistical modeling, which appear under various names, are differentiated (Hand, 2019), namely, explanatory and predictive modeling (Shmueli, 2010). Both types of modeling are concerned with the analysis of data. However, they differ in their purpose and treat data differently.

The purpose of predictive modeling is to provide accurate predictions. The statistical model is used to generate these predictions (Shmueli, 2010). The data at hand are typically split, and one part of the data is used to train the model, while the other part is used to validate the model. Often, the statistical model is treated as a black box (Breiman, 2001), harking back to the remark that, "[i]f a model is created to make this prediction, it should not be constrained by the requirement of interpretability" (Kuhn and Johnson, 2013, p. 4). Consequently, a predictive model does not need to be based on a proper theory but can be driven by data. In contrast, explanatory modeling aims at understanding the mechanisms and processes underlying the data at hand, the so-called data-generating process or population. Explanatory models are typically based on a researcher's theory and are often "simpler" than predictive models, because this facilitates their interpretation (James *et al.*, 2013, Chapter 2.1.1). The data at hand are used for model estimation and testing (causal) hypotheses.

The type of statistical modeling determines how a model is validated. In predictive modeling, model validation focuses on the predictive power of a model, i.e. its ability to accurately predict new/unknown data. In contrast, in explanatory modeling, model validation investigates whether the specified model adequately describes the processes and mechanisms in question (Shmueli, 2010). Therefore, the adequacy of the specified model is of utmost importance because a wrongly specified model likely leads to wrong conclusions. Although in empirical research the line

between predictive and explanatory modeling often may be blurred (e.g. various models from marketing research can be used for predictive purposes, see Leeflang and Wittink, 2000), there are instances where following the rules of explanatory modeling leads to suboptimal solutions in the sense of predictive modeling and vice versa (Ebbes *et al.*, 2011).

SEM is typically regarded as an approach to explanatory modeling (Bollen, 1989; Kline, 2015). Structural equation models are specified in accordance with a theory and estimated to test this theory (Hayduk *et al.*, 2007). To estimate the model parameters, consistent estimators are preferred because a researcher wants to be sure that for a large sample size and a correctly specified model, the estimates are close to the population values with a high probability. This fact is also recognized in the marketing literature: "If the model has a descriptive or normative purpose, consistent estimates are key" (Ebbes *et al.*, 2011, p. 1121). This also highlights the problem of PLS-PM's inconsistency for reflective and causal–formative measurement models comprising latent variables (Dijkstra, 1985).

A crucial step of model validation in SEM is the overall model fit assessment, which means investigating how well the model explains the data (Kline, 2015, p. 120). If the model is an acceptable representation of reality, the data should be consistent with the model and thus with a researcher's theory from which the model is derived. To investigate the overall fit of a model, it is typically examined whether the constraints imposed by the model, which are reflected in the model-implied variance–covariance matrix of the observed variables, are consistent with the collected data, i.e. the sample variance–covariance matrix. It is emphasized that, "if SEM is used, then model fit testing and assessment is paramount, indeed crucial, and cannot be fudged for the sake of 'convenience' or simple intellectual laziness on the part of the investigator" (Barrett, 2007, p. 823). Against this background, the importance of model fit assessment does not depend on a particular estimator. However, different estimators allow for different ways of assessing an estimated model.

The composite model estimated by PLS-PM serves the same purpose as the model known from SEM with latent variables, that is, it represents a researcher's theory. However, they differ in how abstract concepts are represented. While in latent variable models, abstract concepts are represented by latent variables, in composite models, abstract concepts are represented by emergent variables. If composite models are applied in the realm of explanatory modeling, assessing their overall model fit is of the same importance as in SEM with latent variables because it provides an important opportunity to empirically validate a researcher's theory.

While the concept of model fit is well elaborated for structural models containing latent variables that are estimated by maximum likelihood (ML) or related estimators (Hu and Bentler, 1999; Hayduk, 2014), in the context of composite models estimated by PLS-PM, the literature studying the overall model fit assessment is scarce. Hence, in the following section, we adopt methods of the overall model fit assessment from the literature on SEM with latent variables and explain how they can be used to assess the overall fit of composite models estimated by PLS-PM.

4. Ways to assess the overall fit of composite models

In the literature on SEM with latent variables, various methods of assessing the overall model fit have been proposed. They can be broadly categorized into statistical tests and fit indices. Arguably, the most famous test for the overall model fit is the χ^2 test (Jöreskog, 1969). It is based on the asymptotic properties of the fitting function that is minimized by the ML estimator. Since in the context of PLS-PM, no such parametric test has been derived, a nonparametric bootstrap-based alternative was proposed (Dijkstra and Henseler, 2015a; Dijkstra, 2017). Typically, statistical tests for the overall model fit assess the exact model fit,

i.e. the null hypothesis that the specified model is able to exactly reproduce all the variances and covariances among the observed variables in the population.

Although the statistical testing framework is theoretically appealing, testing the exact overall model fit has been criticized as highly unrealistic. The basis of this concern is rooted in Box's (1976) famous remark that "all models are wrong," which implies that the null hypothesis of a perfect fit is always wrong and whether the null hypothesis is rejected is only a matter of sample size. Following this reasoning, the exact model fit is typically not of actual interest to researchers who study a certain phenomenon by means of a model, which is a deliberate approximation of reality (Bentler and Bonett, 1980; Steiger and Lind, 1980).

Against this background, researchers in the early 1980s started popularizing fit indices as an alternative and supplement to the exact model fit testing (Bentler and Bonett, 1980; Jöreskog and Sörbom, 1982). These indices can be roughly categorized into absolute and relative fit indices (McDonald and Ho, 2002). While absolute fit indices measure the correspondence between the specified model and the data along a continuum to gauge how well the model fits (Mulaik *et al.*, 1989), relative fit indices compare the specified model to a reference model to assess the relative increase in model fit (Bentler, 1990). Consequently, fit assessment through fit indices becomes an assessment of *approximate* and *comparative* fit and is of a descriptive, instead of an inferential, nature.

The construction of some of the fit indices, such as the root mean square error of approximation (RMSEA; Steiger and Lind, 1980) and the non-normed fit index (NNFI; Bentler and Bonett, 1980), are directly related to the asymptotic distribution of the χ^2 test statistic, which is derived from the ML estimator. Since such a test statistic with analogous properties has not been derived for PLS-PM, these fit indices are not considered in the following of the paper; instead, we focus on fit indices not tied to a specific estimator, i.e. the root mean square residual (RMR; Jöreskog and Sörbom, 1982), the standardized root mean square residual (SRMR; Bentler, 1995), the normed fit index (NFI; Bentler and Bonett, 1980) and the goodness-of-fit index (GFI; Jöreskog and Sörbom, 1993) and show that they are suitable for assessing composite models.

4.1 Bootstrap-based test for the exact overall model fit

A bootstrap-based test was suggested in the context of composite models to assess the null hypothesis of exact model fit, H_0: $\Sigma(\theta) = \Sigma$ (Dijkstra, 2017). To draw conclusions about the null hypothesis, the discrepancy between the sample variance–covariance matrix S and the estimated model-implied counterpart $\Sigma(\hat{\theta})$ of the observed variables is considered. To measure the discrepancy between these two matrices, several fitting functions $\left(F\left(S, \Sigma(\hat{\theta})\right)\right)$ have been proposed, such as the fitting function of the ML estimator (F_{ML}; Jöreskog, 1970b). Moreover, in the context of PLS-PM and composite models, the geodesic distance (d_G), the squared Euclidean distance and the SRMR have been proposed (Dijkstra, 2017; Schuberth *et al.*, 2018). All these fitting functions have one factor in common, which is that they are equal to zero if the model perfectly fits the data and larger than zero otherwise. However, as suggested by Bollen and Stine (1992), other fit measures such as the NFI can also be used in combination with the bootstrap-based test.

To obtain the reference distribution of a distance function under the null hypothesis, the bootstrap-based test relies on the transformed data set:

$$X^* = XS^{-\frac{1}{2}}\Sigma(\hat{\theta})^{\frac{1}{2}},$$ (5)

where the matrix X contains the original data set, and S and $\Sigma(\hat{\theta})$ are the sample variance–covariance matrix and the estimated variance–covariance matrix implied by the composite model, respectively. The transformation of the original data set X is necessary to mimic a situation where the null hypothesis is true, i.e. the model perfectly fits the data at hand.

For monotonically increasing fit measures such as the previously presented fitting functions, the null hypothesis is rejected for a given significance level α if the value of the fit measure based on the original sample exceeds the $(1 - \alpha)\%$ quantile of the reference distribution. In contrast, for monotonically decreasing fit measures, the null hypothesis is rejected if the value of the fit measure based on the original sample is below the $\alpha\%$ quantile of the reference distribution. In such situations, a researcher has found empirical evidence against the specified model and, following Jöreskog (1969), can conclude that more information can be extracted from the data than is captured by the specified model.

4.2 The (standardized) root mean square residual
The RMR is an absolute fit index proposed by Jöreskog and Sörbom (1982). The residuals are given as the elements of the matrix $S - \Sigma(\hat{\theta})$. Consequently, the RMR shows the root mean square deviation of the sample variance–covariance matrix S from its estimated model-implied counterpart $\Sigma(\hat{\theta})$:

$$\text{RMR} = \sqrt{\frac{2}{K(K+1)} \sum_{i=1}^{K} \sum_{j=1}^{i} \left(s_{ij} - \sigma(\hat{\theta})_{ij} \right)^2}, \tag{6}$$

where s_{ij} and $\sigma(\hat{\theta})_{ij}$ are the elements from the i-th row and the j-th column of the sample variance–covariance matrix and the estimated model-implied variance–covariance matrix, respectively. A disadvantage of the RMR is that its values depend on not only the misfit of the model but also the size of the (co-)variances of the observed variables. Consequently, interpreting the values without taking the scales of the observed variables into account is hardly possible.

To overcome this problem, the SRMR was introduced (Bentler, 1995), which scales the residuals by the standard deviations of the respective observed variables:

$$\text{SRMR} = \sqrt{\frac{2}{K(K+1)} \sum_{i=1}^{K} \sum_{j=1}^{i} \left(\frac{s_{ij} - \sigma(\hat{\theta})_{ij}}{s_{ii}s_{jj}} \right)^2}. \tag{7}$$

Consequently, the SRMR can be roughly interpreted as the average of the absolute value of residual correlations (Pavlov *et al.*, 2021). Since in PLS-PM the observed variables are typically standardized before the analysis, the SRMR equals the RMR.

The RMR and the SRMR are conceptually meaningful for the assessment of composite models. However, the variances and covariances implied by the composite model must be applied. In this case, both the RMR and the SRMR show desirable properties for composite models. For perfectly fitting composite models, i.e. $\Sigma(\hat{\theta}) = S$, both indices show a value of zero. Similarly, for increasing deviations between the empirical and the model-implied variance–covariance matrix, both indices increase, i.e. the larger the misfit, the larger are the values of the (S)RMR. Moreover, if PLS-PM using Mode B is used to estimate the composite

model, the values of the (S)RMR converge in probability to zero if the model is correctly specified.

4.3 The normed fit index

The NFI is a relative fit index that was originally proposed by Bentler and Bonett (1980). It measures the increase in fit relative to the fit of a null model. Although in general various null models are conceivable, in this paper, we focus on the *independence model* as the null model, which assumes that all observed variables are uncorrelated, i.e. that the model-implied variance–covariance matrix of the observed variables equals the diagonal matrix (Bentler and Bonett, 1980, p. 596). Formally, the NFI is defined as follows (Bentler, 1990):

$$\text{NFI} = \frac{F^0 - F^p}{F^0}, \tag{8}$$

where F^0 and F^p are the values of the fitting function for the proposed model and the null model, respectively.

The principle of the NFI can be directly applied to composite models. The NFI represents the improvement in fit of the specified model against the null model as a proportion of the null model, i.e. the relative fit. If the specified composite model fits the data perfectly, i.e. $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = \boldsymbol{S}$, the distance function F^p is equal to zero, and the NFI equals one. In contrast, if the specified model shows the same fit as the null model, the NFI takes a value of 0. The null model typically produces a worse fit than that of the originally specified model because it contains more parameter constraints. Therefore, the NFI ranges from zero to one.

Originally, the ML fitting function was proposed to measure the discrepancy between the sample and the model-implied variance–covariance matrix of the observed variables. In fact, the use of any fitting function that equals zero in the case of perfect fit and is monotonously increasing for increasing misfit is conceivable. This is the case for the discrepancy measures proposed in the context of PLS-PM, i.e. the geodesic distance, the squared Euclidean distance and the SRMR.

4.4 Goodness-of-fit index

The GFI is also a relative fit index (Jöreskog and Sörbom, 1993). It appears to be inspired by the coefficient of determination known from regression analysis and "measures the relative amount of variances and covariances in the empirical covariance matrix $\boldsymbol{S}$ that is predicted by the model-implied covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$" (Schermelleh-Engel *et al.*, 2003, p. 42). The exact definition of the GFI depends on the fitting function used (Mulaik *et al.*, 1989). Only recently has the GFI been proposed in the context of composite models (Cho *et al.*, 2020), when it was defined by means of the unweighted least squares fitting function:

$$\text{GFI} = 1 - \frac{tr\left(\left[\boldsymbol{S} - \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\right]^2\right)}{tr(\boldsymbol{S}^2)}, \tag{9}$$

where tr denotes the trace operator, and $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{S}$ indicate the estimated variance–covariance matrix implied by the composite model and the empirical counterpart, respectively. Consequently, the GFI equals 1 if the composite model perfectly fits the data set, i.e. when $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = \boldsymbol{S}$, and values below 1 indicate a misfit.

5. Concerns about the overall model fit assessment in the context of PLS-PM

In the context of PLS-PM, several concerns regarding the overall model fit assessment have been raised. The following subsections discuss these concerns and provide a conclusion.

5.1 Concern 1: PLS-PM is focused on estimating causal–predictive relationships; hence, model fit assessment is of little value

The literature argues that PLS-PM was developed as an approach to causal–predictive modeling (Wold, 1982), and consequently, model fit assessment is of little value (Hair *et al.*, 2020).

Unfortunately, neither Wold (1982), the founder of PLS-PM, nor the literature that refers to his work provides a clear definition of causal–predictive modeling. Hence, it can have different meanings. First, and following Hair *et al.* (2019a) and recent literature that aims at demystifying the role of causal–predictive modeling (Chin *et al.*, 2020), causal–predictive modeling could be a middle way between explanatory and predictive modeling that strives to achieve the goals of both explanatory and predictive modeling. Although this idea is striking, it can hardly be achieved, as "the 'wrong' model can sometimes predict better than the correct one" (Shmueli, 2010, p. 293), and explanatory power does not imply predictive power (Forster and Sober, 1994). Against this background, it is not clear why model fit assessment should be disregarded following this understanding of causal–predictive modeling. Second, causal–predictive modeling can mean that researchers make use of explanatory models to make predictions, i.e. model-based predictions. Compared to approaches known from predictive modeling, such as artificial neural networks (Haykin, 2009), this approach has the advantage of knowing how the predictions were made because explanatory models are usually "simple" to ensure their interpretability. On the other hand, it is likely that this approach is inferior to predictive models which are not tied to an explanatory model. Since this approach is based on an explanatory model, it is not clear why researchers should not rely on common principles of explanatory modeling, such as the overall model fit assessment in the context of SEM, to discard wrongly specified models. Similarly, Lohmöller (1989, p. 73) notes that "the predictive purpose should not jeopardize a structural-causal interpretation of the relation." It is well known that correctly specified models may exhibit high out-of-sample predictive accuracy, but reversing the argument is a logical fallacy (Saylors and Trafimow, 2020). In fact, several studies have shown that researchers relying on the measurement evaluation steps of PLS-SEM, which replace the overall model fit assessment with predictive measures, miss an important opportunity to detect wrongly specified models (McIntosh *et al.*, 2014; Schuberth, 2020). Hence, replacing the overall model fit assessment with predictive measures is not recommended for researchers conducting explanatory modeling, regardless of whether predictions are subsequently made.

Conclusion: Since causal–predictive modeling is not clearly defined in the PLS-PM literature, researchers should not use it as a justification to omit the step of the overall model fit assessment when they are working (at least partially) in the realm of explanatory modeling, i.e. testing theories and drawing statistical inferences.

5.2 Concern 2: Assessing the model fit by means of distance measures makes no sense in the context of PLS-PM because PLS-PM does not minimize these distances

The PLS-PM literature is concerned about the overall model fit assessment by means of distance functions that measure the discrepancy between the estimated model-implied and the sample variance–covariance matrix of the observed variables (Hair *et al.*, 2019b). Similar concerns have been raised about the bootstrap-based test for the overall model fit, which is

based on a distance function and "should be considered with extreme caution" (Hair *et al.*, 2019a, p. 31). These concerns are rooted in the fact that PLS-PM does not minimize such a distance function, in contrast to the ML estimator, to obtain the parameter estimates.

In fact, PLS-PM does not minimize a distance function to obtain the parameter estimates but iteratively estimates several regressions by ordinary least squares. However, as shown by Dijkstra (2017), PLS-PM produces consistent estimates for the composite model like the ML estimator does for common factor models (Jöreskog, 1970b). Moreover, both the ML estimator and PLS-PM are Fisher consistent for common factor and composite models, respectively, and are asymptotically normal (Dijkstra, 2010). Consequently, PLS-PM shows similar statistical properties for composite models as the ML estimator shows for common factor models, although they obtain their estimates differently.

To assess the overall model fit by means of a distance function, it is reasonable to assume a consistent estimator because it produces a consistently estimated model-implied variance–covariance matrix. Otherwise, a distance function would indicate a model misfit even when the sample size converges to infinity, which is of course not desirable. The way in which the estimates are produced plays only a minor role as long as they are consistent. In contrast, the specified model is of much greater importance because an estimator loses its statistical properties, such as consistency, if applied to the wrong model. Hence, it is not clear why model fit assessment by means of a distance function should only function for models that have been estimated by an estimator that minimizes that distance function. The SEM literature has already provided examples of model fit assessment by means of a distance function in cases where an estimator was used that does not minimize such a distance function (Devlieger *et al.*, 2019). Similarly, the SRMR, which can be regarded as a type of distance function, is often considered to assess common factor models that have been estimated by ML; the ML estimator does not minimize the SRMR. This provides additional support against the claim that the overall model fit assessment by means of a distance function makes no sense if the estimator does not minimize this distance function.

In general, quantifying the misfit between the estimated model-implied and sample variance–covariance matrix can be done by any function that accepts these two matrices as input. However, for interpretational purposes, it is desirable that these functions have some particular properties. First, they should be equal to zero if the two matrices are identical, i.e. zero should indicate a perfect fit. Second, they should monotonically increase with an increasing difference between the two matrices. Meeting these requirements, a larger value of the distance function indicates a larger misfit of the model. It is noted that the ML fitting function, the SRMR, the squared Euclidean distance and the geodesic distance function meet these requirements. However, at this stage, the threshold values up to which the discrepancy in the model fit is regarded as acceptable remain unclear (see Subsection 5.4).

To assess the exact model fit via statistical significance testing, one needs to have the (asymptotic) distribution of the distance function under the null hypothesis, i.e. that the model-implied variance–covariance matrix based on the population parameters equals the population variance–covariance matrix of the observed variables $\left(H_0 : \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}\right)$. This distribution depends on several aspects, such as the distance function used and the distributions of the estimated model-implied and the sample variance–covariance matrix. For example, it is well known that the number of observations minus one times the ML fitting function based on the ML estimates given multivariate normally distributed observed variables asymptotically follows a χ^2 distribution under the null hypothesis of exact fit (Jöreskog, 1970a). In contrast, for distance functions based on PLS-PM estimates, such a distribution is generally not known.

To overcome the distributional assumptions, a bootstrap-based test was developed that can be used to assess the null hypothesis of exact fit (Beran and Srivastava, 1985, and

Section 4.1). Although this test was first proposed to assess structural equation models containing latent variables (Bollen and Stine, 1992), it can be applied in the same way to assess structural models containing emergent variables (Dijkstra, 2017). It does not require an estimator that minimizes a specific distance but rather a consistently estimated model-implied variance–covariance matrix (Beran and Srivastava, 1985) given by an estimator that produces consistent parameter estimates [4]. This is the case for PLS-PM using Mode B if applied to estimate composite models (Dijkstra, 2017).

Conclusion: Distance functions and the bootstrap-based test can be used to assess the overall fit of composite models, even though PLS-PM does not minimize such a distance function.

5.3 Concern 3: Fit indices that are based on the common factor model are not appropriate to assess a model estimated by PLS-PM

The PLS-PM literature became concerned quite early about the use of fit indices based on a common factor model to assess models estimated by PLS-PM (Lohmöller, 1989, p. 54).

It is generally not appropriate to estimate common factor models by PLS-PM because it produces inconsistent estimates for this type of model (Dijkstra, 1985). Consequently, evaluating the fit of a common factor model estimated by PLS-PM is not recommended because even for a correctly specified model and a sample that converges to infinity, fit indices would indicate a misfit. Researchers who want to apply PLS-PM to estimate common factor models should instead use consistent partial least squares and its enhancements (Dijkstra and Henseler, 2015a,b; Rademaker *et al.*, 2019), which provide consistent estimates for common factor models.

Typically, fit indices are based on the model-implied variance–covariance matrix, which captures the constraints imposed by the underlying model. As shown in Section 2, not only the common factor model but also the composite model impose such constraints. These constraints can be exploited in fit indices to assess the overall fit of composite models. Obviously, it is important to apply the variance–covariance matrix implied by the composite model (see also Section 4).

Conclusion: The fit of the composite model can be assessed by fit indices proposed in the context of SEM with latent variables if the variance–covariance matrix implied by the common factor model is replaced by the one implied by the composite model.

5.4 Concern 4: Thresholds for fit indices have not been proposed for composite models

The PLS-PM literature reveals concerns about the fact that no threshold values for fit indices have been proposed for composite models (Hair *et al.*, 2019a). As a consequence, it is difficult for researchers applying PLS-PM to judge the absolute and relative fit of their models.

The SEM literature has suggested various threshold values for fit indices that can be applied to judge common factor models (Hu and Bentler, 1999), while in the context of composite models, only a single study proposes such thresholds (Cho *et al.*, 2020). Although we think that fit indices are helpful to quantify the degree of model misfit, we are skeptical about comparing the value of a fit index to a threshold value derived from simulation studies to decide whether the model fit is acceptable. As highlighted in the SEM literature, this approach is problematic in several ways: First, it is very difficult, if not even impossible, to generalize such thresholds beyond the simulation design because the distribution of fit indices is influenced by factors other than the degree of misspecification that fit indices attempt to quantify (Yuan, 2005). Consequently, deriving and proposing threshold values is of little benefit for applied researchers, whose research setting likely differs from the design

of the simulation study. Second, deriving threshold values through a simulation is based on a flawed logic because the degree of misfit that is still regarded as acceptable is determined by the simulation designer in advance (Marsh *et al.*, 2004). Hence, it falls to the subjective judgment of the simulation designer to determine which model fits are acceptable or unacceptable. Third, deriving threshold values for fit indices under the hypothesis of exact fit contradicts the logic underlying absolute fit indices. Although SEM literature began to embed fit indices in a testing framework (Bollen and Stine, 1992), absolute fit indices were originally introduced to overcome the issue of exact fit. Hence, determining threshold values as quantiles of the distribution of a fit index under perfect fit contradicts the very concept of approximate fit. In fact, it was shown in the context of SEM with latent variables that the conventional χ^2 test outperforms the index-plus-threshold-value decision strategy in distinguishing correctly from incorrectly specified models (Marsh *et al.*, 2004).

Conclusion: The use of fit indices is a controversial topic in the literature on SEM with latent variables. The concerns can generally be transferred to the composite model. While opponents call for abandoning the use of fit indices (Barrett, 2007), there are also more optimistic voices that regard fit indices as useful tools to assess the overall model fit. For example, fit indices can be beneficial in situations where the sample size is large and the test for exact fit rejects the null hypothesis, although it is only trivially false (Bentler, 2007). Hence, we take a more liberal stand and recommend reporting fit indices along with the results of the test for exact model fit, because they can provide additional information. However, we recommend against the common practice of comparing fit indices to threshold values derived by simulation studies to judge whether the fit of a composite model is acceptable because this approach suffers from logical inconsistency (Marsh *et al.*, 2004).

5.5 Concern 5: PLS-PM is used in case of small sample sizes for which small misspecification is not reliably detected by the bootstrap-based test

The literature argues that PLS-PM is often used in case of sample sizes for which the bootstrap-based test shows only low statistical power, i.e. misspecification often remains undetected (Hair *et al.*, 2020). Hence, its use is of only little value in the context of PLS-PM.

It is well known that the power of a statistical test decreases with decreasing sample size as the sampling uncertainty increases (Cohen, 1988, Chapter 1). Hence, this behavior is not an idiosyncrasy of the bootstrap-based test but applies to all statistical significance tests.

Small sample sizes are particularly concerning in the context of explanatory modeling, and the importance of sufficiently large sample sizes has already been highlighted in the context of SEM (Kline, 2015) and marketing research (Sawyer and Ball, 1981). Hence, researchers using PLS-PM who are working in the realm of explanatory modeling are advised to collect a sufficient amount of data before conducting their analysis. As recognized by Rigdon (2016, p. 600), one could say that "PLS path modeling will produce parameter estimates even when [the] sample size is very small, but reviewers and editors can be expected to question the value of those estimates, beyond simple data description."

To address this issue, researchers using SEM with latent variables are usually advised to investigate a priori whether the size of the collected sample is sufficiently large to ensure that the statistical test being used has sufficient power, e.g. by conducting Monte Carlo simulations (Wolf *et al.*, 2013). The same approach is also recommended in the context of PLS-PM (Aguirre-Urreta and Rönkkö, 2015). In principle, similar guidelines can be followed to assess the statistical power of the bootstrap-based test of the overall fit of composite models. However, such guidelines have not yet been elaborated.

Conclusion: Like all statistical significance tests, the power of the bootstrap-based test for the overall model fit depends on the sample size. If analysts deem the statistical power too

low, they should collect more data. To not test a model is the worst option: It corresponds to a statistical power of zero.

5.6 Concern 6: It is not clear whether fit should be assessed based on the estimated model or a model with a saturated structural model
The PLS-PM literature raises concerns about which model should actually be assessed, i.e. the estimated model or the model with a saturated structural model (Hair *et al.*, 2019b).

Recent PLS-PM guidelines for explanatory modeling recommend first assessing the composite model with a saturated structural model and subsequently assessing the originally specified model (Henseler *et al.*, 2016; Benitez *et al.*, 2020); see Section 7 for a more elaborate presentation. The idea of this approach is rooted in the two-step procedure that has been proposed in the context of SEM with latent variables (Anderson and Gerbing, 1988). Among applied researchers, this approach is regarded as beneficial because it allows us to localize the source of misfit, i.e. whether the composition of the emergent variables (first step) *or* the complete model (second step) is problematic. Ultimately, it is the originally specified model that represents a researcher's theory, and therefore, its fit is what needs to be assessed.

Conclusion: Analysts should assess the fit of their originally specified model. Assessing the fit of a model with a saturated structural model can serve as a useful intermediate step in model fit assessment to localize potential sources of misfit.

6. Monte Carlo simulation
An important and still open question is the efficacy of the bootstrap-based test for the overall fit and the various fit measures presented, i.e. the geodesic distance, the SRMR, the NFI and the GFI. To answer this question, we conduct a Monte Carlo simulation. Since the comparison of fit indices to derived threshold values has been strongly criticized in the SEM literature (Marsh *et al.*, 2004), we deliberately do not aim at deriving any threshold values for these fit measures but instead investigate their finite-sample performance in combination with the bootstrap-based test for the exact overall model fit. In particular, we examine the type I error rate and the statistical power of the bootstrap-based test.

We consider three scenarios comprising three different population models. Each population model consists of three emergent variables. The three scenarios, including their population models, parameters and variance-covariance matrices, are displayed in Figure 1. Since the bootstrap-based test was recently evaluated with regard to wrongly specified relationships between observed variables and emergent variables (Schuberth *et al.*, 2018), we exclusively focus on misspecifications in the structural model. Therefore, in all population models, only the structural model differs across the scenarios, whereas the weights and the intra-block correlation matrices are kept constant.

Scenario 1 is considered to assess the test's type I error rate. In this scenario, the estimated model matches the population model, and thus, the estimated model is correctly specified. As shown in Figure 1, the SRMR and the geodesic distance are equal to zero if they are calculated for the estimated model based on the population variance–covariance matrix. Similarly, the NFI and GFI show a value of 1. For this scenario, we expect that the bootstrap-based test produces rejection rates close to the predefined significance level.

Scenarios 2 and 3 serve to assess the statistical power of the bootstrap-based test. In Scenario 2, the estimated model does not match the population model, i.e. the estimated model is misspecified. As seen in Figure 1, in the population model of Scenario 2, there is a direct effect between the emergent variables η_1 and η_3 that is omitted in the estimated model. Consequently, the SRMR and the geodesic distance show values larger than 0 for the estimated model based on the population variance–covariance matrix. Similarly, the NFI
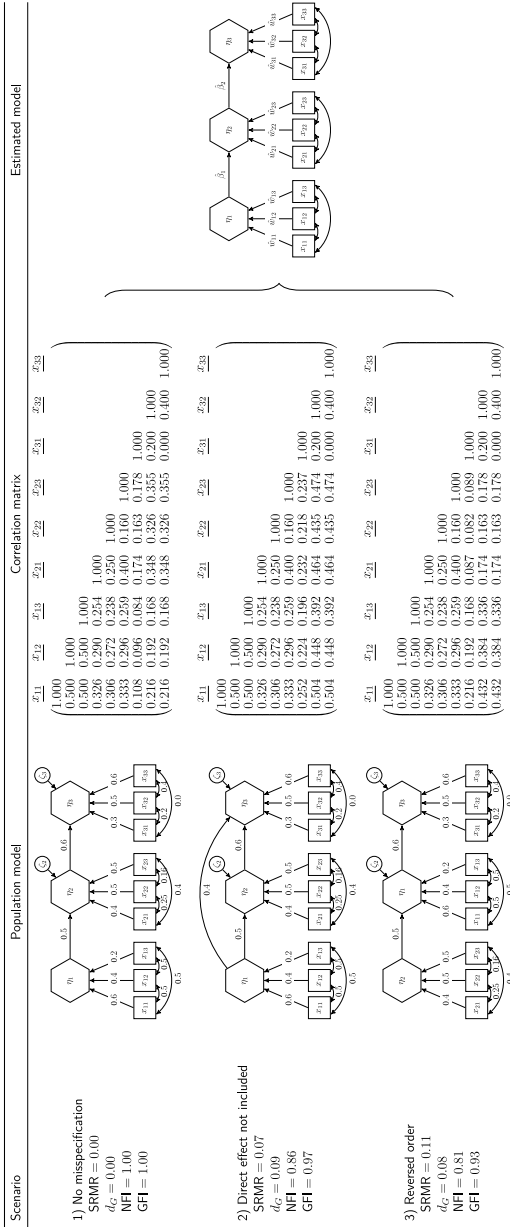
Notes: The presented values of the fit measures refer to the estimated model based on the population correlation matrix of the observed variables. The values in the correlation matrices and the fit indices are rounded to the third and second decimal, respectively

and GFI values are 0.86 and 0.97, respectively, in this scenario. Therefore, we expect that the bootstrap-based test for the overall model fit produces rejection rates above the predefined significance level.

In the population model of Scenario 3, the role of the emergent variables η_1 and η_2 in the structural model is switched in comparison to that in the estimated model. Consequently, the estimated model is misspecified. As shown in Figure 1, the SRMR and the geodesic distance are 0.11 and 0.08, respectively, and highlight a misfit of the estimated model based on the population variance–covariance matrix. Similarly, the GFI and NFI values are smaller than 1. Against this background, we expect that the bootstrap-based test produces rejection rates above the predefined significance level.

It is noteworthy that the different fit measures assess the two misspecifications differently. As shown in Figure 1, the SRMR, the NFI and the GFI indicate a worse model fit for the model in Scenario 3, whereas the geodesic distance indicates a worse fit for the model in Scenario 2. We expect that this will also be reflected in the test's rejection rates.

To study the finite-sample behavior of the bootstrap-based test, we vary the sample size from 50, 100, 250, 500, and 1,000 to 2,000 observations per sample. Moreover, we consider two significance levels, namely, 1% and 5%. As is common, for larger sample sizes, we expect an increase in the statistical test's power when the estimated model is indeed misspecified. Similarly, we expect higher statistical power in the case of a higher significance level.

The complete simulation was conducted in the statistical programming environment R (R Core Team, 2020). For each condition, 1,000 samples were drawn from a multivariate normal distribution with mean zero and the variance–covariance matrix of the respective scenario using the mvrnorm function of the MASS package (Venables and Ripley, 2002). To estimate the specified model by PLS-PM, the csem function of the cSEM package was used (Rademaker and Schuberth, 2020). For the inner weighting, the factorial weighting scheme was used, and for the estimation of the weights, Mode B was applied. As a stopping criterion, the absolute change in the weights was considered. If the largest absolute difference was smaller than 10^{-5}, the algorithm would stop. Furthermore, the maximum number of iterations was set to 1,000. To run the bootstrap-based test for the overall model fit, the testOMF function of the cSEM package was used. Although we did not face any convergence issues for the initial PLS-PM estimations, we replaced estimations that may not have converged during the bootstrap to ensure that all bootstrap-based tests are based on 499 valid bootstrap runs.

Figure 2 illustrates the results of our simulation. For Scenario 1, i.e. the scenario in which the estimated model is correctly specified, the test produces rejection rates slightly below the predefined significance level for small sample sizes, i.e. $n \leq 100$, regardless of the assumed significance level and the fit measure used. However, for an increasing sample size, the produced rejection rates converge toward the assumed significance level.

Considering Scenarios 2 and 3, i.e. in the case of model misspecification, the rejection rates are below the recommended threshold of 80% (Cohen, 1988) for very small sample sizes, i.e. $n = 50$, regardless of the fit measure used. However, in line with our expectations, the produced rejection rates increase for an increasing sample size, and for sample sizes larger than 100 observations, the produced rejection rates were above 80%. Moreover, the rejection rates are higher for the larger significance levels, confirming our expectations. Comparing the performance of the fit measures across Scenarios 2 and 3, the results are largely in line with our expectations. The bootstrap-based test in combination with the geodesic distance rejects the model in Scenario 2 more often than the model in Scenario 3, while the test based on the SRMR and the GFI detect the
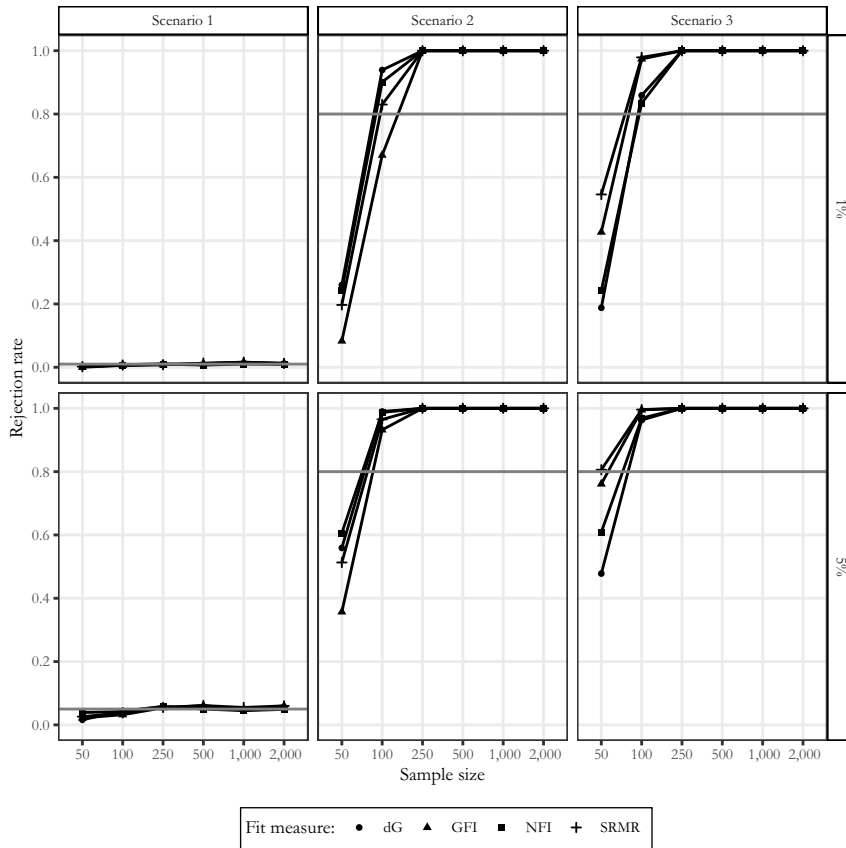
Figure 2.
Rejection rates for
Scenarios 1, 2, and 3

misspecification of Scenario 3 more reliably. Considering the bootstrap-based test in combination with the NFI, the results are not that clear, i.e. in some conditions, it rejects the model from Scenario 2 more often, while in other conditions, it rejects the model from Scenario 3 more often. We would have expected it to reject the model from Scenario 3 more often.

To conclude, the bootstrap-based test for the overall model fit in combination with the presented fit measures is able to detect model misspecification and produces rejection rates close to the predefined significance levels when the estimated model is correctly specified. However, a sufficient sample size is required to achieve satisfactory statistical power. For all considered fit measures, the bootstrap-based test behaved as expected, i.e. the rejection rates increased for an increasing sample size and/or larger significance levels when the estimated model was misspecified. However, the sensitivity of the studied fit measures for model misspecification differs with regard to the kind of misspecification. The geodesic distance indicates a larger misfit for the model in Scenario 2 than for the model in Scenario 3, while the SRMR and GFI show the opposite. For the NFI, the picture is not that clear.

7. Guidelines on the assessment of the overall fit of composite models

Figure 3 depicts our guidelines to assess the overall fit of composite models estimated by PLS-PM. To eventually assess the overall fit of composite models including a structural model, we recommend a two-step procedure known from current guidelines on the use of PLS-PM in confirmatory and explanatory research (Benitez *et al.*, 2020). In the first step, a CCA is conducted, while in the second step, the fit of the originally specified model is assessed. This way of model fit assessment is recommended, as it is a logical necessity that the abstract concepts be properly operationalized before the analysis of the structural model is performed (Anderson and Gerbing, 1982). To illustrate the approach, we focus on a researcher who derived from her theory the model displayed in Figure 4.

In the first step, a CCA is conducted, i.e. a model in which the emergent variables freely correlate is estimated, and its overall fit is assessed. Typically, the originally specified model is nested in this model, i.e. the originally specified model contains more restrictions on the parameters than the model with freely correlated emergent variables. Figure 5 displays the model for our researcher from the first step. This model exhibits the same fit as the originally specified model from Figure 4 with a saturated structural model.

An unsatisfactory fit in the first step indicates that the operationalization of the abstract concepts as emergent variables should be reconsidered, as the emergent variables do not convey all the information between the observed variables from two different blocks. Consequently, there are problems in the composition of at least one emergent variable. In contrast, if the fit of the model in the first step is satisfactory, the researcher can continue with the second step.

In the second step, the originally specified model (the model from Figure 4) is estimated and assessed. If the model does not show an acceptable fit, likely the structural model is misspecified. For our fictitious researcher, this can mean that the emergent variable η_2 does not fully meditate the effect of η_1 on η_3.

The advantage of the two-step approach in comparison to a one-step approach is that a researcher can better localize the source of misfit. In case misfit is detected, regardless in which step, the researcher is advised to inspect its source. For example, a researcher can follow guidelines known from the SEM literature (Kline, 2015) and investigate the residuals. Moreover, in reporting the model fit assessment results, we recommend providing the outcomes of the criteria mentioned in Section 4.

8. Discussion

The overall model fit assessment in the context of SEM is crucial if SEM is applied for explanatory modeling. Its importance is widely acknowledged in the SEM literature, although not without controversies, e.g. the discourse in the special issue of the journal *Personality and Individual Differences* (Vernon and Eysenck, 2007). In contrast, for composite models estimated by PLS-PM, it is less clear when and how to assess the overall model fit. To address these issues, we explain that the overall fit assessment of composite models is of utmost importance if composite models are studied in the context of explanatory modeling. Thus, the role of the overall fit assessment is unaffected by the way that abstract concepts are modeled, i.e. as latent or emergent variables. Moreover, we present a bootstrap-based test and four fit indices and show that they are all suitable for assessing the overall fit of composite models estimated by PLS-PM.

The PLS-PM literature has raised several concerns about model fit assessment and its applicability when PLS-PM is used for model estimation (Lohmöller, 1989; Hair *et al.*, 2017, 2019b,a, 2020). The present study discusses these concerns and shows that most of them are unfounded. The current understanding of causal–predictive modeling does not warrant omission of the overall model fit assessment if researchers use the composite model and
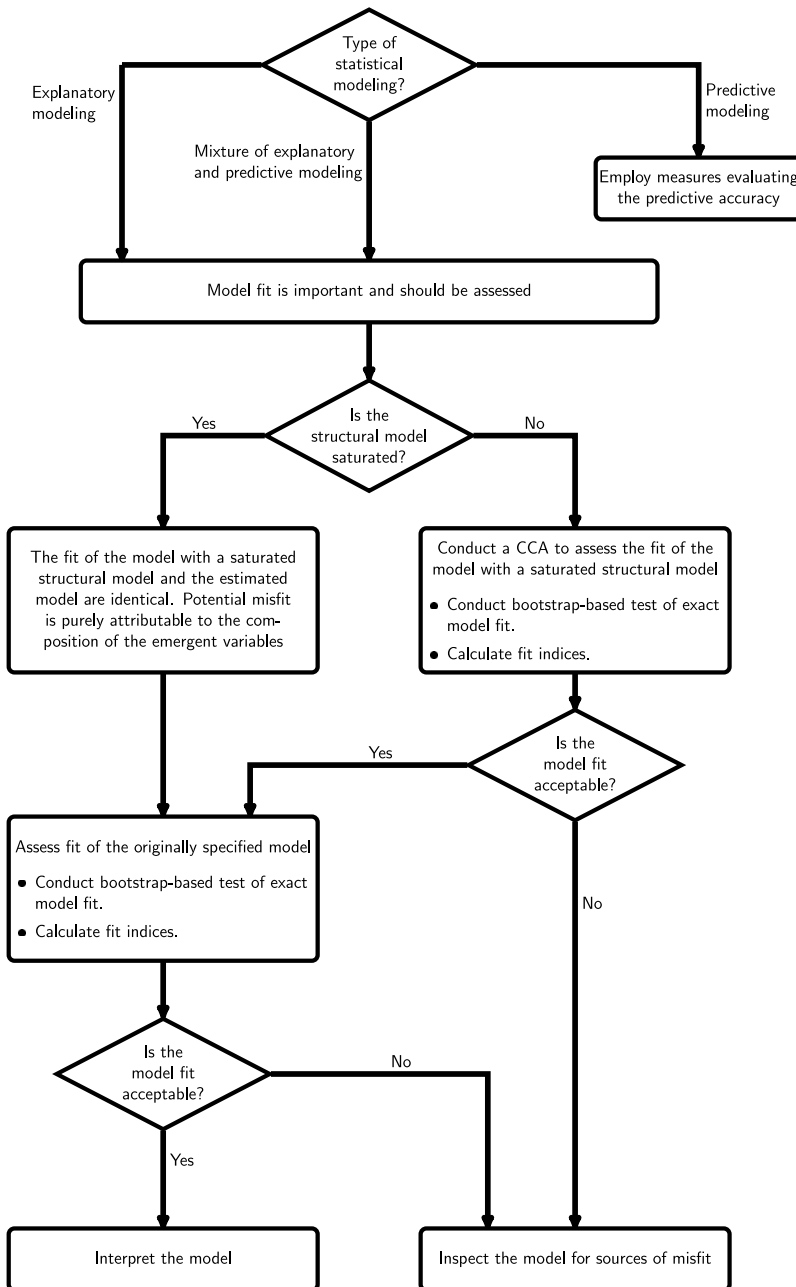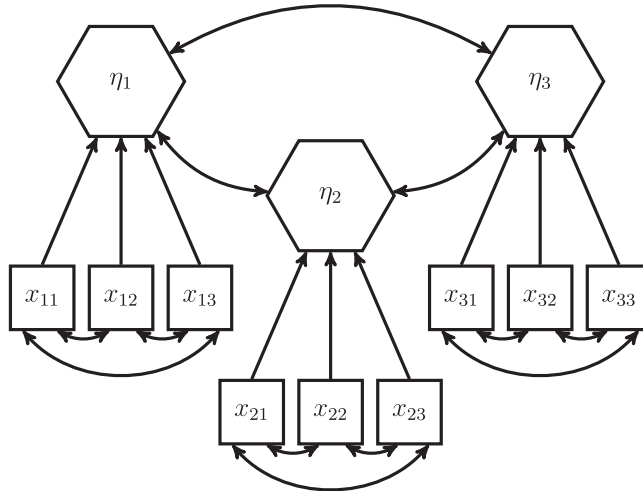
Figure 3.
Guidelines on
assessing the overall
fit of composite
models

Figure 4.
Originally specified
composite model
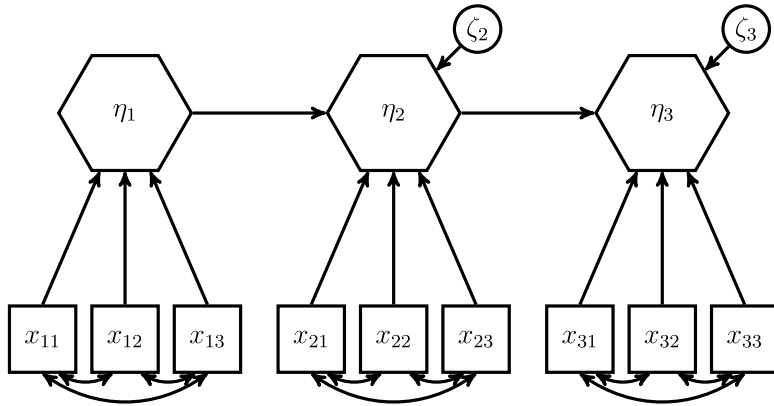
Figure 5.
CCA model from the
first step

PLS-PM for theory testing. If PLS-PM is used in the context of explanatory modeling, the overall model fit assessment is a pivotal step. Moreover, the use of the overall model fit criteria that are based on the model-implied variance–covariance matrix is appropriate to assess composite models even though these criteria were first developed for common factor models. However, the variance–covariance matrix implied by the composite model must be applied. Similarly, composite models estimated by PLS-PM can be assessed by means of distance functions even though PLS-PM does not minimize such a function to obtain the parameter estimates. The bootstrap-based test for the overall model fit can be used to assess the exact fit of a composite model. As shown by our simulation, it is able to detect misspecified models estimated by PLS-PM in finite samples and it can also be used in combination with fit indices such as the NFI and the GFI. Although its statistical power might be insufficiently low owing to small sample sizes, this is no reason to abandon the bootstrap-based test. However, it is important that researchers are aware of that risk. Finally, fit indices can quantify the approximate and relative fit of composite models,

although it is not recommended to judge the fit of a model based on threshold values derived from simulation studies.

To support researchers applying PLS-PM in the overall fit assessment of their composite models, the current study provides concise guidelines, i.e. a two-step assessment procedure. While in the first step, a CCA is conducted, in the second step, the originally specified model is assessed. This approach helps researchers better localize the source of misfit. For each step, we recommend reporting the results of the bootstrap-based test and the values of the SRMR, the NFI and the GFI. It is emphasized that researchers who act in the realm of explanatory modeling should take model fit assessment seriously, otherwise they will miss an important opportunity for model validation. Guidelines on PLS-PM for explanatory research that discourage the assessment of model fit resemble cooking recipes that suggest a visual and haptic inspection but at the same time discourage tasting the meal.

Our study is limited to the bootstrap-based test for the overall model fit and fit indices that have been proposed to assess the overall fit of composite models. In general, other ways have been suggested to assess composite models, including prediction tests, prediction metrics, tests for rank restrictions on submatrices and the exploitation of differences between different estimators (Dijkstra, 2017; Shmueli *et al.*, 2019; Liengaard *et al.*, 2020). However, none of these should replace the overall model fit assessment in the context of explanatory modeling. Moreover, we limit our focus on the (S)RMR, the NFI and the GFI, as the principles of these fit indices are not tied to the asymptotic properties of a specific estimator. Although we have shown that in principle, the NFI and the GFI can detect misspecified composite models, the SEM literature has shown that they are affected by the sample size and model complexity (Hu and Bentler, 1998, 1999; Sharma *et al.*, 2005). Therefore, alternatives such as the NNFI (Bentler and Bonett, 1980) have been proposed. In this regard, we recommend investigating whether the principles underlying the NNFI, and similarly the RMSEA, also apply to composite models estimated by PLS-PM. Furthermore, our simulation study showed that the fit measures assess misspecifications differently. Therefore, future research should identify situations in which a specific fit measure is preferred.

Similarly, our proposed guidelines are limited to linear and recursive models estimated by PLS-PM. The limitation to PLS-PM is in no way mandatory, and other estimators that produce consistent estimates for composite models are valid alternatives. Moreover, in empirical research, scientists encounter situations where models are non-recursive, e.g. the models contain feedback loops (Dijkstra, 2017). It is noted that the presented guidelines can still be applied for this type of model. Additionally, non-recursive models often provide the opportunity to exploit the involved overidentification restriction through statistical tests such as the Sargan–Hansen test (Sargan, 1958) to investigate whether the postulated assumptions required for identification hold.

In SEM, the issue of equivalent models is well known in the literature (Raykov and Penev, 1999) and often encountered in empirical research (MacCallum *et al.*, 1993). Equivalent models exhibit identical levels of model fit, i.e. they all produce the same model-implied variance–covariance matrix even when the model parameter estimates differ (Raykov and Penev, 1999). Consequently, the overall model fit assessment cannot help identify the *correct* model among all equivalent models. It is obvious that the issue of equivalent models is not specific to latent variable models but also applies to composite models. Hence, to validate a model, a researcher needs to argue why his/her model should not be rejected in favor of an equivalent model (Kline, 2015).

Notes

1. The notion of an emergent variable is used to emphasize that the composite conveys all the information between its antecedents and its consequences and that it is on the same level as a latent variable. Moreover, emergent variables composed of latent variables, emergent variables, or a mixture of both are conceivable (Van Riel *et al.*, 2017; Schuberth *et al.*, 2020). However, in this article, we focus on emergent variables made up of observed variables.

2. Only recently, it was shown that a special type of composite model in which the emergent variables are composed of correlation weights can be consistently estimated by PLS-PM Mode A (Cho and Choi, 2020). This type of composite model is a special case of the composite model presented by Dijkstra (2017) and Schuberth *et al.* (2018) and can also be consistently estimated by PLS-PM Mode B.

3. It is emphasized that we do not refer to the measurement evaluation steps known from PLS-SEM, which have also been recently dubbed as confirmatory composite analysis (Hair *et al.*, 2020). For a comparison of the two, we refer to Schuberth (2020).

4. In addition, the bootstrap-based test for overall model fit is based on commonly made assumptions such as independent and identically distributed (i.i.d.) observed variables (Beran and Srivastava, 1985).

References

Aguirre-Urreta, M. and Rönkkö, M. (2015), "Sample size determination and statistical power analysis in PLS using R: an annotated tutorial", *Communications of the Association for Information Systems*, Vol. 36, pp. 33-51.

Anderson, J.C. and Gerbing, D.W. (1982), "Some methods for respecifying measurement models to obtain unidimensional construct measurement", *Journal of Marketing Research*, Vol. 19 No. 4, pp. 453-460.

Anderson, J.C. and Gerbing, D.W. (1988), "Structural equation modeling in practice: a review and recommended two-step approach", *Psychological Bulletin*, Vol. 103 No. 3, pp. 411-423.

Barrett, P. (2007), "Structural equation modelling: adjudging model fit", *Personality and Individual Differences*, Vol. 42 No. 5, pp. 815-824.

Benitez, J., Henseler, J., Castillo, A. and Schuberth, F. (2020), "How to perform and report an impactful analysis using partial least squares: guidelines for confirmatory and explanatory is research", *Information and Management*, Vol. 2 No. 57, p. 103168.

Bentler, P.M. (1990), "Comparative fit indexes in structural models", *Psychological Bulletin*, Vol. 107 No. 2, pp. 238-246.

Bentler, P.M. (1995), *EQS Structural Equations Program Manual*, Vol. 6, Multivariate Software, Inc., Encino, CA.

Bentler, P.M. (2007), "On tests and indices for evaluating structural models", *Personality and Individual Differences*, Vol. 42 No. 5, pp. 825-829.

Bentler, P.M. and Bonett, D.G. (1980), "Significance tests and goodness of fit in the analysis of covariance structures", *Psychological Bulletin*, Vol. 88 No. 3, pp. 588-606.

Beran, R. and Srivastava, M.S. (1985), "Bootstrap tests and confidence regions for functions of a covariance matrix", *The Annals of Statistics*, Vol. 13 No. 1, pp. 95-115.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, John Wiley and Sons, New York, NY.

Bollen, K.A. and Stine, R.A. (1992), "Bootstrapping goodness-of-fit measures in structural equation models", *Sociological Methods and Research*, Vol. 21 No. 2, pp. 205-229.

Box, G.E.P. (1976), "Science and statistics", *Journal of the American Statistical Association*, Vol. 71 No. 356, pp. 791-799.

Breiman, L. (2001), "Statistical modeling: the two cultures", *Statistical Science*, Vol. 14 No. 3, pp. 190-231.

Chin, W., Cheah, J.-H., Liu, Y., Ting, H., Lim, X.-J. and Cham, T.H. (2020), "Demystifying the role of causal-predictive modeling using partial least squares structural equation modeling in information systems research", *Industrial Management and Data Systems*, Vol. 120 No. 12, pp. 2161-2209.

Cho, G. and 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*, Vol. 47 No. 1, pp. 243-272.

Cho, G., Hwang, H., Sarstedt, M. and Ringle, C.M. (2020), "Cutoff criteria for overall model fit indexes in generalized structured component analysis", *Journal of Marketing Analytics*, Vol. 8 No. 4, pp. 189-202.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Devlieger, I., Talloen, W. and Rosseel, Y. (2019), "New developments in factor score regression: fit indices and a model comparison test", *Educational and Psychological Measurement*, Vol. 79 No. 6, pp. 1017-1037.

Dijkstra, T.K. (1985), *Latent Variables in Linear Stochastic Models: Reflections on 'Maximum Likelihood' and 'Partial Least Squares' Methods*, Vol. 1, Sociometric Research Foundation, Amsterdam.

Dijkstra, T.K. (2010), "Latent variables and indices: Herman Wold's basic design and partial least squares", in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds), *Handbook of Partial Least Squares*, Springer, Berlin, Heidelberg, pp. 23-46.

Dijkstra, T.K. (2017), "A perfect match between a model and a mode", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, Springer, Cham, pp. 55-80.

Dijkstra, T.K. and Henseler, J. (2015a), "Consistent and asymptotically normal PLS estimators for linear structural equations", *Computational Statistics and Data Analysis*, Vol. 81 No. 1, pp. 10-23.

Dijkstra, T.K. and Henseler, J. (2015b), "Consistent partial least squares path modeling", *MIS Quarterly*, Vol. 39 No. 2, pp. 297-316.

Ebbes, P., Papies, D. and van Heerde, H.J. (2011), "The sense and non-sense of holdout sample validation in the presence of endogeneity", *Marketing Science*, Vol. 30 No. 6, pp. 1115-1122.

Forster, M. and Sober, E. (1994), "How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions", *The British Journal for the Philosophy of Science*, Vol. 45 No. 1, pp. 1-35.

Hair, J.F., Howard, M.C. and Nitzl, C. (2020), "Assessing measurement model quality in PLS-SEM using confirmatory composite analysis", *Journal of Business Research*, Vol. 109, pp. 101-110.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2017), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, London.

Hair, J.F., Sarstedt, M., Ringle, C.M. and Mena, J.A. (2012), "An assessment of the use of partial least squares structural equation modeling in marketing research", *Journal of the Academy of Marketing Science*, Vol. 40 No. 3, pp. 414-433.

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019a), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24.

Hair, J.F., Sarstedt, M. and Ringle, C.M. (2019b), "Rethinking some of the rethinking of partial least squares", *European Journal of Marketing*, Vol. 53 No. 4, pp. 566-584.

Hand, D. (2019), "What is the purpose of statistical modelling?", *Harvard Data Science Review*, Vol. 1 No. 1, pp. 1-6.

Hayduk, L. (2014), "Seeing perfectly fitting factor models that are causally misspecified understanding that close-fitting models can be worse", *Educational and Psychological Measurement*, Vol. 74 No. 6, pp. 905-926.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H. and Boulianne, S. (2007), "Testing! Testing! One, two, three – testing the theory in structural equation models!", *Personality and Individual Differences*, Vol. 42 No. 5, pp. 841-850.

Haykin, S. (2009), *Neural Networks and Learning Machines*, 3rd ed, Pearson, New York, NY.

Henseler, J. (2017), "Bridging design and behavioral research with variance-based structural equation modeling", *Journal of Advertising*, Vol. 46 No. 1, pp. 178-192.

Henseler, J. (2018), "Partial least squares path modeling: Quo vadis?", *Quality and Quantity*, Vol. 52 No. 1, pp. 1-8.

Henseler, J. (2021), *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables*, Guilford Press, New York, NY.

Henseler, J. and Dijkstra, T. (2017), *Adanco 2.0.1, Kleve, Germany: Composite Modeling*.

Henseler, J. and Schuberth, F. (2020), "Using confirmatory composite analysis to assess emergent variables in business research", *Journal of Business Research*, Vol. 120, pp. 147-156.

Henseler, J., Hubona, G. and Ray, P.A. (2016), "Using PLS path modeling in new technology research: updated guidelines", *Industrial Management and Data Systems*, Vol. 116 No. 1, pp. 2-20.

Hu, L-T. and Bentler, P.M. (1998), "Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification", *Psychological Methods*, Vol. 3 No. 4, pp. 424-453.

Hu, L-T. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 6 No. 1, pp. 1-55.

Hubona, G.S., Schuberth, F. and Henseler, J. (2021), "A clarification of confirmatory composite analysis (CCA)", *International Journal of Information Management*, Vol. 61, p. 102399, doi: 10.1016/j.ijinfomgt.2021.102399.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013), *An Introduction to Statistical Learning*, 7th ed, Springer, New York, NY.

Jöreskog, K. (1969), "A general approach to confirmatory maximum likelihood factor analysis", *Psychometrika*, Vol. 34 No. 2, pp. 183-202.

Jöreskog, K.G. (1970a), "A general method for analysis of covariance structures", *Biometrika*, Vol. 57 No. 2, pp. 239-251.

Jöreskog, K.G. (1970b), "A general method for estimating a linear structural equation system", *ETS Research Bulletin Series*, Vol. 1970 No. 2, p. 41.

Jöreskog, K.G. and Sörbom, D. (1982), "Recent developments in structural equation modeling", *Journal of Marketing Research*, Vol. 19 No. 4, pp. 404-416.

Jöreskog, K.G. and Sörbom, D. (1993), *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*, Scientific Software International Lawrence Erlbaum, Chicago, Ill. Hillsdale, NJ.

Khan, G.F., Sarstedt, M., Shiau, W.-L., Hair, J.F., Ringle, C.M. and Fritze, M.P. (2019), "Methodological research on partial least squares structural equation modeling (PLS-SEM): an analysis based on social network approaches", *Internet Research*, Vol. 29 No. 3, pp. 407-429.

Kline, R.B. (2015), *Principles and Practice of Structural Equation Modeling*, Guilford Press, 4th ed, New York, NY, London.

Kuhn, M. and Johnson, K. (2013), *Applied Predictive Modeling*, Springer, New York, NY.

Leeflang, P.S. and Wittink, D.R. (2000), "Building models for marketing decisions", *International Journal of Research in Marketing*, Vol. 17 Nos 2/3, pp. 105-126.

Leek, J.T. and Peng, R.D. (2015), "What is the question?", *Science*, Vol. 347 No. 6228, pp. 1314-1315.

Liengaard, B.D., Sharma, P.N., Hult, G.T.M., Jensen, M.B., Sarstedt, M., Hair, J.F. and Ringle, C.M. (2020), "Prediction: coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling", *Decision Sciences*, Vol. 52 No. 2.

Lohmöller, J.-B. (1989), *Latent Variable Path Modeling with Partial Least Squares*, Physica, Heidelberg.

MacCallum, R.C., Wegener, D.T., Uchino, B.N. and Fabrigar, L.R. (1993), "The problem of equivalent models in applications of covariance structure analysis", *Psychological Bulletin*, Vol. 114 No. 1, pp. 185-199.

Marsh, H.W., Hau, K.-T. and Wen, Z. (2004), "In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 11 No. 3, pp. 320-341.

McDonald, R.P. and Ho, M.-H.R. (2002), "Principles and practice in reporting structural equation analyses", *Psychological Methods*, Vol. 7 No. 1, pp. 64-82.

McIntosh, C.N., Edwards, J.R. and Antonakis, J. (2014), "Reflections on partial least squares path modeling", *Organizational Research Methods*, Vol. 17 No. 2, pp. 210-251.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S. and Stilwell, C.D. (1989), "Evaluation of goodness-of-fit indices for structural equation models", *Psychological Bulletin*, Vol. 105 No. 3, pp. 430-445.

Pavlov, G., Maydeu-Olivares, A. and Shi, D. (2021), "Using the standardized root mean squared residual (SRMR) to assess exact fit in structural equation models", *Educational and Psychological Measurement*, Vol. 81 No. 1, pp. 110-130.

R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rademaker, M.E. and Schuberth, F. (2020), "cSEM: composite-based structural equation modeling", R package version: 0.3.0.1.9000, available at: https://m-e-rademaker.github.io/cSEM/

Rademaker, M.E., Schuberth, F. and Dijkstra, T.K. (2019), "Measurement error correlation within blocks of indicators in consistent partial least squares", *Internet Research*, Vol. 29 No. 3, pp. 448-463.

Raykov, T. and Penev, S. (1999), "On structural equation model equivalence", *Multivariate Behavioral Research*, Vol. 34 No. 2, pp. 199-244.

Rigdon, E.E. (2016), "Choosing PLS path modeling as analytical method in European management research: a realist perspective", *European Management Journal*, Vol. 34 No. 6, pp. 598-605.

Ringle, C., Wende, S. and Becker, J.-M. (2015), *SmartPLS 3*, Boenningstedt: SmartPLS GmbH.

Sargan, J.D. (1958), "The estimation of economic relationships using instrumental variables", *Econometrica*, Vol. 26 No. 3, pp. 393-415.

Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O. and Gudergan, S.P. (2016), "Estimation issue with PLS and CBSEM: where the bias lies!", *Journal of Business Research*, Vol. 69 No. 10, pp. 3998-4010.

Sawyer, A.G. and Ball, A.D. (1981), "Statistical power and effect size in marketing research", *Journal of Marketing Research*, Vol. 18 No. 3, pp. 275-290.

Saylors, R. and Trafimow, D. (2020), "Why the increasing use of complex causal models is a problem: on the danger sophisticated theoretical narratives pose to truth", *Organizational Research Methods*, Vol. 24 No. 3.

Schermelleh-Engel, K., Moosbrugger, H. and Müller, H. (2003), "Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures", *Methods of Psychological Research Online*, Vol. 8 No. 2, pp. 23-74.

Schuberth, F. (2020), "Confirmatory composite analysis using partial least squares: setting the record straight", *Review of Managerial Science*, Vol. 15, pp. 1311-1345, doi: 10.1007/s11846-020-00405-0.

Schuberth, F., Henseler, J. and Dijkstra, T.K. (2018), "Confirmatory composite analysis", *Frontiers in Psychology*, Vol. 9, p. 2541.

Schuberth, F., Rademaker, M.E. and Henseler, J. (2020), "Estimating and assessing second-order constructs using PLS-PM: the case of composites of composites", *Industrial Management and Data Systems*, Vol. 120 No. 12, pp. 2211-2241.

Sharma, S., Mukherjee, S., Kumar, A. and Dillon, W.R. (2005), "A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models", *Journal of Business Research*, Vol. 58 No. 7, pp. 935-943.

Shmueli, G. (2010), "To explain or to predict?", *Statistical Science*, Vol. 25 No. 3, pp. 289-310.

Shmueli, G., Sarstedt, M., Hair, J.F., Cheah, J.-H., Ting, H., Vaithilingam, S. and Ringle, C.M. (2019), "Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict", *European Journal of Marketing*, Vol. 53 No. 11, pp. 2322-2347.

Steiger, J.H. and Lind, J.C. (1980), "Statistically-based tests for the number of common factors", in Paper presented at the annual meeting of the Psychometric Society, IA City, IA.

Van Riel, A.C.R., Henseler, J., Kemény, I. and Sasovova, Z. (2017), "Estimating hierarchical constructs using partial least squares: the case of second order composites of factors", *Industrial Management and Data Systems*, Vol. 117 No. 3, pp. 459-477.

Venables, W.N. and Ripley, B.D. (2002), *Modern Applied Statistics with S*, Springer.

Vernon, T. and Eysenck, S. (2007), "Introduction", *Personality and Individual Differences*, Vol. 42 No. 5, p. 813.

Wold, H. (1982), "Soft modeling: the basic design and some extensions", in Jöreskog, K.G. and Wold, H., (Eds), *Systems under Indirect Observation: Causality – Structure – Prediction Part II, Chap. 1*, North-Holland Publishing Company, Amsterdam, pp. 1-54.

Wolf, E.J., Harrington, K.M., Clark, S.L. and Miller, M.W. (2013), "Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety", *Educational and Psychological Measurement*, Vol. 73 No. 6, pp. 913-934.

Yuan, K.-H. (2005), "Fit indices versus test statistics", *Multivariate Behavioral Research*, Vol. 40 No. 1, pp. 115-148.

Corresponding author
Jörg Henseler can be contacted at: j.henseler@utwente.nl