Reflections on Partial Least Squares Path Modeling

Running Head: Reflections on PLS-PM

Cameron N. McIntosh*

National Crime Prevention Centre

Public Safety Canada

Ottawa, Ontario, Canada

Jeffrey R. Edwards

Kenan-Flagler Business School

University of North Carolina

Chapel Hill, NC

John Antonakis

Faculty of Business and Economics

University of Lausanne

Lausanne-Dorigny, Switzerland

***Corresponding author**:
Cameron N. McIntosh, National Crime Prevention Centre, Public Safety Canada, 269 Laurier Avenue West, K1A 0P8 Ottawa, Ontario, Canada.
Email: cnm100@hotmail.com

Note: None of the abovementioned authors have any commercial interests in any of the statistical software packages discussed in this article.

Abstract

The purpose of the present article is to take stock of a recent exchange in *Organizational Research Methods* between critics (Rönkkö & Evermann, 2013) and proponents (Henseler et al., 2014) of partial least squares path modeling (PLS-PM). The two target articles were centered around six principal issues, namely whether PLS-PM: (1) can be truly characterized as a technique for structural equation modeling (SEM); (2) is able to correct for measurement error; (3) can be used to validate measurement models; (4) accommodates small sample sizes; (5) is able to provide null hypothesis tests for path coefficients; and (6) can be employed in an exploratory, model-building fashion. We summarize and elaborate further on the key arguments underlying the exchange, drawing from the broader methodological and statistical literature in order to offer additional thoughts concerning the utility of PLS-PM and ways in which the technique might be improved. We conclude with recommendations as to whether and how PLS-PM serves as a viable contender to SEM approaches for estimating and evaluating theoretical models.

Reflections on Partial Least Squares Path Modeling

Introduction

Partial least squares path modeling (PLS-PM) has begun to achieve widespread usage among applied researchers. Starting with the initial work by Wold (1966, 1973, 1975), the application of PLS-PM has been stimulated by comprehensive expositions and computer implementations by Lohmöller (1984, 1987, 1988, 1989), Chin (1988, 1998, 2003) and others (for detailed historical reviews of the development of PLS-PM, see Mateos-Aparicio, 2011; Trujillo, 2009). PLS-PM has also received thorough treatment in a number of textbooks (Abdi, Chin, Vinzi, Russolillo, & Trinchera, 2013; Hair, Hult, Ringle, & Sarstedt, 2014; Vinzi, Chin, Henseler, & Wang, 2010), and both proprietary and open source software packages for conducting PLS-PM are now widely available (Addinsoft, 2013; Chin, 2003; Kock, 2013; Monecke, 2013; Ringle, Wende, & Will, 2005; Rönkkö, 2013; Sanchez & Trinchera, 2013). PLS-PM is gaining a particularly strong foothold in fields such as marketing and information systems research, as evidenced by three special journal issues during the past three years: one in the *Journal of Marketing Theory and Practice* (Hair, Ringle, & Sarstedt, 2011) and two in *Long Range Planning* (Hair, Ringle, & Sarstedt, 2012, 2013). PLS-PM has also spread to the organizational sciences (Antonakis, Bastardoz, Liu, & Schriesheim, 2014; Hulland, 1999; Rönkkö & Everman, 2013), and its momentum appears to be on the rise.

PLS-PM has garnered a wide following largely due to beliefs by its users that it has important advantages over other analytical techniques, such as regression analysis, structural equation modeling (SEM), and simultaneous equation estimators (e.g., two-stage and three-stage least squares). However, methodological discussions of PLS-PM have raised questions about its statistical underpinnings and its viability as an estimation procedure. For instance, a number of reviewers have found that PLS-PM practitioners do not fully acknowledge its various pitfalls and have offered detailed methodological guidelines intended to remedy or avoid these pitfalls (e.g., Hair, Sarstedt, Ringle, & Mena, 2012; Hair et al., 2013; Henseler, Ringle, & Sinkovics, 2009; Hair, Sarstedt, Pieper, & Ringle, 2012; Marcoulides & Chin, 2013; Marcoulides & Saunders, 2006; Peng & Lai, 2012; Ringle, Sarstedt, & Straub, 2012). Other critics go further, maintaining that regardless of how rigorously PLS-PM is applied, it suffers from intractable statistical flaws that warrant a drastic reduction of its use (Goodhue, Thompson, & Lewis, 2013), or even its complete abandonment (e.g., Antonakis, Bendahan, Jacquart, Lalive, 2010; Rönkkö, in press; Ronkko & Everman, 2013; Rönkkö & Ylitalo, 2010). A common theme of these critiques is that the availability of proven, powerful, and versatile modeling techniques, such as SEM, can preclude the use of PLS-PM altogether. In response to these mounting concerns, some proponents of PLS-PM have devised innovative statistical approaches to improve its performance in both model estimation and testing (e.g., Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press). Whereas the preliminary theoretical and empirical evidence for these new strategies appears promising, these methods are still in their infancy and have yet to be fully evaluated through a comprehensive program of simulation research.

A recent manifestation of the tensions between the critics and proponents of PLS-PM has appeared in two articles published in *Organizational Research Methods.* From the critical perspective, Rönkkö and Evermann (2013) used a series of conceptual arguments and empirical demonstrations in an attempt to show that several commonly held beliefs about particular properties and capabilities of PLS-PM – namely, that it is in fact an SEM technique, is able to correct for measurement error, can validate measurement models, works well in small samples, is able to provide null hypothesis tests on path coefficients, and can be used in an exploratory, model-building fashion – are "methodological myths and urban legends" (p. 426). In response, Henseler et al. (2014) provided a point-by-point rebuttal to each of Rönkkö and Evermann's critiques, maintaining that most of their arguments are based on narrowly-conceived simulations and fundamental misconceptions about the purposes and capabilities of PLS-PM. Whereas the exchange presented by these articles raises issues that are timely and important, the potential usefulness of this debate for guiding future work on PLS-PM is hampered by the fact that the issues raised were left largely unresolved. Moreover, the exchange did not capture some additional issues that are relevant to the potential utility of PLS-PM, which draw from the broader literature on multivariate data analysis.

The purpose of the present article is to (a) summarize and reflect on the key issues that underlie the exchange between Rönkkö-Evermann and Henseler et al.; (b) attempt to resolve these issues in an even-handed manner; and (c) draw from other areas of statistical theory and practice (e.g., psychometrics, econometrics, SEM, causal graphs) to offer additional thoughts concerning the utility of PLS-PM and ways in which the PLS-PM estimator might be improved. We hope to provide food for thought that will interest both critics and proponents of PLS-PM, with the intent of challenging both sides to seek common ground concerning the overriding goal of applied statistical modeling, which is to provide unbiased and efficient estimates of model parameters that allow meaningful tests of theories that embody important substantive phenomena. We conclude with recommendations as to whether and how PLS-PM serves as a viable contender to SEM approaches in model estimation and evaluation. Table 1 summarizes the positions on PLS-PM across all three articles, with respect to each of the six core issues and an overall judgment on whether PLS-PM should be abandoned as a statistical tool for applied research.

Insert Table 1 about here

----

1. Can PLS-PM be characterized as an SEM method?

Rönkkö and Evermann began their critique by questioning the treatment of PLS-PM as a method of SEM, based on two arguments. First, PLS-PM estimates path models not with latent variables, but with weighted linear composites of observed variables. Second, rather than using path analysis with simultaneous equations, PLS uses separate OLS regressions that estimate relationships between the composites. For these reasons, Rönkkö and Evermann asserted that PLS is more akin to OLS regressions on summed scales or factor scores than to covariance structure analysis. The authors subsequently moderated their position, however, saying that "although PLS can technically be argued to be an SEM estimator, so can OLS regression with summed scales or factor scores: Both fit the definition of the term *estimator* (emphasis in original) (Lehmann & Casella, 1998, p. 4) because they provide some estimates of model parameters" (p. 433). Nonetheless, Rönkkö and Evermann continued by criticizing the quality of PLS-PM estimates, noting that they are biased and inconsistent, and further added that "the lack of an overidentification test is another disadvantage of PLS over SEM" (p. 433).

Henseler et al. countered by citing previous researchers, including Wold, who have characterized PLS-PM as a method of SEM. Henseler et al. also argued that PLS-PM falls well within various descriptions of SEM, which generally refer to statistical techniques that examine relationships between independent and dependent variables within a presumed causal structure (e.g., Byrne, 1998; Ullman & Bentler, 2001). In addition, they maintained that the core, underlying statistical framework for PLS-PM is the "composite factor model", a more general case of the common factor model that is the foundation of SEM. Moreover, Henseler et al. criticized SEM for its reliance on the common factor model, arguing that this model "*rarely* holds in applied research" (emphasis in original) and that other models should be considered, particularly the composite factor model ascribed to PLS-PM. Henseler et al. noted that, because this model places no restrictions on the covariances among the items assigned to a factor, it will generally yield better fit than the common factor model. Further, they pointed out that PLS-PM estimates are only biased and inconsistent if they are viewed as estimates of common rather than composite factor model parameters. Given that the two models are conceptually and statistically distinct, the estimates need to be interpreted differently. Finally, Henseler et al. noted that a global overidentification test can be used in PLS-PM to verify the causal specification of the model, just as in the SEM context.

In our view, arguing whether PLS-PM should be called an SEM method obscures the primary substance of the debate. The more important issues concern the type of modeling approach represented by PLS-PM and its adequacy for estimating and testing hypothesized causal structures, regardless of whether this approach is characterized as SEM. Therefore, we will focus on the methodological specifics of the approach and set aside the controversy regarding the labeling of PLS-PM as a SEM method.

First, we note that Henseler et al.'s formal statistical distinction between the composite and common factor models stands in stark contrast to the traditional PLS-PM canon. PLS-PM was originally developed as a less computationally demanding alternative to maximum likelihood-based SEM for estimating the associations among latent variables (e.g., Jöreskog &

Wold, 1982; Wold, 1982, 1985), not as a method for estimating structural relations among composite variables. Because the composites in PLS-PM contain measurement error, the technique has historically been treated as simply a convenient and rough approximation of the common factor model, only capable of producing consistent estimates of factor loadings and intercorrelations as both the sample size (N) and number of observed indicators (p) increase without bound (i.e., *consistency at large*; Haenlein & Kaplan, 2004). However, it seems that Henseler et al. are aligned with Rigdon's (2012) recent recommendations to free PLS-PM from its original grounding in common factor-based SEM and develop it further as a purely composite-based modeling approach, thereby precluding any undue comparisons with true latent variable models.

Although we see some merit in this perspective, Henseler et al.'s presentation of the "composite factor model" is problematic in several respects. In particular, the path diagram that Henseler et al. use to illustrate their composite factor model is actually a common factor model containing within-factor correlated measurement errors. As presented, the model is not identified, meaning that no unique solution exists for the parameters (Davis, 1993). Although this identification problem could be resolved by imposing additional parameter constraints (e.g., setting the measurement error correlations to be equal) or using informative Bayesian prior distributions (Lee & Song, 2012; McIntosh, 2013; Muthén & Asparouhov, 2012), a more serious issue is that the model's causal structure and parameterization are fundamentally inconsistent with the model that PLS-PM (and other component-based approaches) actually estimates. To illustrate, we contrast Henseler et al.'s representation of the composite factor model with a more accurate depiction of a composite-based measurement model in Figure 1, where: (a) the measurement-level pathways lead from the indicators to the composites, consistent with the manner in which composite variables are formed in PLS-PM models; (b) the indicators are specified as having no measurement error; and (c) the composites also have no error, meaning that they are exact weighted linear combinations of their indicators (Bollen, 2011; Bollen & Bauldry, 2011; Grace & Bollen, 2008; Kline, 2013a). This composite measurement model can be estimated using PLS-PM, other component-based modeling techniques (e.g., Hwang & Takane, 2004; Hwang, 2008, 2009; Tenenhaus, 2013; Tenenhaus & Tenenhaus, 2011), or even SEM provided that modified parameterizations are used (Bollen, 2011, Bollen & Bauldry, 2011; Bollen & Davis, 2009; Dolan, 1996; Dolan, Bechger, & Molenaar, 1999; McDonald, 1996; Treiblmaier, Bentler, & Mair, 2011).

Given these marked differences between the two models, Henseler et al.'s characterization of the composite factor model as a more general case of the common factor model is not tenable. In order for this nesting relation to hold, the more restricted model must be derivable from the less restricted model by imposing constraints on parameters (e.g., fixing parameters to some constant, such as zero, or setting parameters equal to one another; Steiger, Shapiro, & Browne, 1985). In other words, the constraints in the less restricted model must be a strict subset of those in the more restricted model. However, one cannot generate a common factor model simply by fixing parameters in the composite factor model displayed in our Figure 1, as the direction of causality in the measurement model is reversed (i.e., indicator→construct rather than construct→indicator), and measurement error variances and covariances do not exist in the composite case. Thus, Henseler et al.'s exposition would have been better served by focusing on PLS-PM as a technique for estimating strictly component-based path models (Tenenhaus, 2008), rather than trying to recast PLS-PM in terms of the common factor model.

Despite Henseler et al.'s critique of the common factor model, a broader survey of the literature reveals that PLS-PM proponents have not fully declared independence from the factor-analytic tradition. For instance, a method called consistent PLS (PLSc) has been devised that allows PLS-PM to recover common factor model parameters in finite samples (Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press). Briefly, PLSc compensates for the absence of a true measurement model by using a rescaling method to disattenuate factor loadings and intercorrelations for measurement error. Dijkstra (2014) argues that PLSc allows PLS-PM to accommodate both composite and common factor models, noting that an exclusive focus on composite-based modeling would substantially limit the usefulness of the technique. Although PLSc is an impressive development, it is questionable whether PLSc adds any value over common factor-based SEM's more versatile and powerful estimation and testing procedures, an issue we later address in greater detail.

Turning now to model testing, we agree with Henseler et al. that the recent development of a global chi-square fit statistic for PLS-PM represents an important theoretical and empirical advance (Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press). Prior to this development, the evaluation of PLS-PM models consisted of examining measures of explained variance, which shed little light on the tenability of a causal model (Henseler & Sarstedt, 2013). To be sure, explained variance is an important element of model quality, as it quantifies the strength of hypothesized relationships and the potential impact of interventions. However, because the accuracy of parameter estimates hinges on achieving acceptable fit, measures of variance explained are subordinate to tests of fit (Antonakis et al., 2010; Hayduk, Pazderka-Robinson, Cummings, Levers, & Beres, 2005; McIntosh, 2007). Thus, model fit should be established prior to evaluating model parameters, including measures of explained variance.

Several additional issues relevant to testing PLS-PM models merit attention. First, a global chi-square statistic merely provides an omnibus test of all constrained parameters (e.g., hypothesized null pathways) in the target model (Jöreskog, 1969). Therefore, a significant chi-square statistic does not identify which particular aspects of the model are at odds with the observed data. For this reason, omnibus tests of model fit should be supplemented with local tests of fit on individual constraints to identify the specific sources of model misspecification (Bera & Bilias, 2001; Saris, Satorra, & van der Veld, 2009), as well as an assessment of which estimated pathways are most affected by the misspecifications (Kolenikov, 2011; Yuan, Kouros, & Kelley, 2008; Yuan, Marshall, & Bentler, 2003). Such procedures are currently available when using maximum likelihood-based techniques. However, PLS-PM lags far behind in this area of model evaluation, thus limiting the extent to which the researcher can verify fit and ensure the interpretability of parameter estimates. Indeed, there are many ways in which a PLS-PM model could show lack of fit. For instance, typically not all of the composites in a given model will be linked by direct causal pathways, such that some of the inner relations will be indirect. Further, because all of the between-block information is assumed to be conveyed by the composites, observed variables from one block are assumed to have no direct connections with those from other blocks.

One approach that can be used to conduct local tests of PLS-PM models involves the vanishing partial correlations implied by the model (Elwert, 2013; Hayduk et al., 2003; Shipley, 2000, 2003, Pearl, 2009). To illustrate, consider the basic mediational model: $A \rightarrow B \rightarrow C$, which implies that A and C are *conditionally independent* given B; more formally, $A \perp C \mid B$. For this model, a test of the partial correlation $r_{AC.B}$ indicates whether full mediation (i.e., a zero direct effect of A on C) holds, a procedure known in econometrics as a Hausman test (Hausman, 1978, 1983; see also Abrevaya, Hausman, & Khan, 2010; Antonakis et al., 2010; Antonakis, Bendahan, Jacquart, & Lalive, 2014). This method is feasible for PLS-PM models, because it merely requires the correlation matrix of the composites and observed variables from an unrestricted measurement model and the relevant partial correlations implied by the model, with inferences performed using bootstrapping techniques. Because identifying all implied conditional independencies can be tedious and prone to error, software for causal graphs should be used for this purpose (Kyono, 2010; Marchetti, Drton, & Sadeghi, 2013; Textor, 2013).

In addition to fixed zero parameters in PLS-PM models, *equality* constraints on certain free parameters may be required to evaluate hypothesized differences in the magnitudes of effects. For example, a researcher using PLS-PM could have a theory predicting that the structural pathways from two explanatory composite variables to a composite outcome variable are of different strengths. In addition, one might want to determine whether the measurement-level relationships between individual indicators and composite variables are invariant across certain types of population sub-groupings (e.g., gender, ethnicity). If chi-square difference tests show that model fit significantly deteriorates following the imposition of an equality constraint, then there is evidence that the parameters in question are reliably different from each other (Steiger et al., 1985; Yuan & Bentler, 2004). Unfortunately, PLS-PM software does not currently allow for the imposition of equality constraints (Tenenhaus, 2008; Tenenhaus, Mauger, & Guinot, 2010). However, an alternative option is to use SEM software to mimic the PLS-PM parameterization (e.g., McDonald, 1996), thereby permitting the use of equality constraints within a composite-based path model and conventional chi-square difference tests for evaluating the tenability of the constraints.

Second, Henseler et al.'s response to the Rönkkö and Evermann critique did not explicitly address the concerns regarding endogeneity, which refers to a violation of the key causal modeling assumption that the independent variables in an equation are uncorrelated with the error term, that is, $r_{x,\varepsilon} = 0$ for all x (Antonakis et al., 2010, 2014; Bollen, 2012; McIntosh, 2012; Semadeni, Withers, & Certo, 2013). Unfortunately, endogeneity tends to be the rule rather than the exception in applications of multiple regression and related methods when using observational rather than experimental data. This problem can stem from various factors, such as omitted variables, unmodeled measurement error, selection bias, common method effects, and non-recursive pathways among constructs (e.g., feedback loops). Given that endogeneity causes parameter estimates to be inconsistent, corrective procedures are needed. In econometrics and other areas of applied research, the technique of choice for dealing with endogeneity is instrumental variable estimation (IVE; Angrist and Krueger 2001; Greenland 2000). To counteract problems created by endogeneity, instrumental variables must be: (1) strongly correlated with the independent variables; and (2) independent of the error terms. IVE is typically implemented using two stage least squares (2SLS), with the first stage involving the regression of an independent variable x on an instrument z, followed by computing the predicted values from this equation, as follows:

$$x = \gamma_0 + \gamma_1 z + \mu \tag{1}$$

$$\hat{x} = \hat{\gamma}_0 + \hat{\gamma}_1 z \tag{2}$$

In the second stage, the predicted values are substituted for the original independent variable in the focal explanatory equation:

$$y = \beta_0 + \beta_1 \hat{x} + \varepsilon \tag{3}$$

Because *z* is exogenous, $r_{\hat{x},\varepsilon} = 0$, and β_1 can be estimated consistently. Additional tests are then conducted to verify that the 2SLS estimates differ from those obtained the conventional OLS approach (Abrevaya et al., 2010), and that the instruments are uncorrelated with the error term (Baum, Schaffer, & Stillman, 2003, 2007; Semykina 2012). For the latter type of test to be viable, the model must be overidentified, that is, the number of instruments must be greater than the number of endogenous predictors. Furthermore, 2SLS is not the only approach for implementing IVE, as one can also use simultaneous equation methods such as ML (Antonakis, et al., 2010; Baum et al., 2003, 2007) and three-stage least squares (3SLS) regression (Johnson, Ayinde, & Oyejola, 2011; Belsley, 1992; Kontoghiorghes & Dinenis, 1997).

To our knowledge, however, only two studies have explicitly addressed endogeneity in composite predictors of PLS-PM models (Lovaglio & Vittadini, 2013; Vittadini, Minotti, Fattore, & Lovaglio, 2007). In these studies, Vittadini and his colleagues pointed out that correlations between predictors and outcomes in the explanatory equations of the PLS-PM inner model can be influenced by unmodeled components in the predictor blocks, that is, systematic variation that is orthogonal to the composites of primary theoretical interest. By extracting these extra components from the predictor blocks and explicitly including them in the model, endogeneity bias can be removed (Lovaglio & Vittadini, 2013; Vittadini et al., 2007). However, this approach relies on information that is already available in the predictor blocks and therefore cannot adjust for endogeneity stemming from omitted variables or selection bias. The PLSc approach is similarly limited, as it merely applies a rescaling correction for measurement error to obtain consistent estimates of common factor model parameters (Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press). Therefore, the more comprehensive IVE approach is required to cover the potential causes of endogeneity bias in real applications.

It is noteworthy that IVE has been gaining prominence in the SEM domain, due mainly to the work of Bollen and his colleagues (Bollen & Bauer 2004; Bollen, Kirby, Curran, Paxton, & Chen, 2007; Bollen & Maydeu-Olivares, 2007; Kirby & Bollen, 2009; see also Nestler, 2013a, 2013b). Given the promising results of this work, the potential transportability of IVE methods to the PLS-PM context should be examined in future research. Indeed, given that a 2SLS approach that does not involve instruments is already used in PLSc (Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press), a logical next step is to include instruments to counteract the bias created by endogeneity.

Third, neither Rönkkö and Evermann nor Henseler et al. addressed the issue of correlated errors in the regression equations of the inner model and how these correlations might impact parameter estimates and model fit. In the OLS context, a large body of theoretical and empirical work has addressed seemingly unrelated regression equations (SURE; Zellner, 1963), in which a set of regression equations estimated separately are interrelated via their error terms (Beasley, 2008; Foschi, Belsley, & Kontoghiorghes, 2003; Kubáček, 2013). More generally, when using system-wide estimators (e.g., ML or 3SLS) with simultaneous equation models, error terms should be allowed to correlate given the possibility of omitted causes. To illustrate, assume the following true causal model:

$$y = \beta_0 + \beta_1 x + \beta_2 q + \varepsilon \tag{4}$$

$$x = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 q + \mu \tag{5}$$

Now, suppose that q is omitted, and the model that is actually estimated is:

$$y = \delta_0 + \delta_1 x + \theta \tag{6}$$

$$x = \lambda_0 + \lambda_1 z_1 + \lambda_2 z_2 + \psi \quad (7)$$

Because omitting q means that equations 6 and 7 are misspecified, the parameters differ from those in equations 4 and 5. The absence of q can be accounted for by including $r_{\theta,\psi}$ in the model (Antonakis et al., 2010), after which the true parameter values can then be recovered (i.e., $\delta_1 = \beta_1$, $\lambda_1 = \gamma_1$, etc.). Thus, if errors are not permitted to correlate when they should (cf. Cole, Ciesla, & Steiger, 2007; Reddy, 1992), overidentification tests will fail and coefficient estimates will be inconsistent. Currently, PLS-PM models do not conventionally allow the error terms of the inner model to be intercorrelated, as each equation is estimated independently, typically using the OLS procedure. This problem likely gives rise to additional inconsistency in the coefficients and could be resolved by estimating the entire set of inner relations using system-wide estimators like ML or 3SLS regression (Johnson et al., 2010), which would explicitly take the error correlations into account while also allowing the use of instruments.

2. Can PLS-PM reduce the impact of measurement error?

Rönkkö and Evermann questioned the notion that PLS-PM reduces the effects of measurement error. Their arguments focused on the comparison of the weighted composites involved in PLS-PM with the unweighted sums of items used in OLS regression, noting that any advantage of PLS-PM must derive from the weights assigned to the indicators, which is essentially the only feature that sets PLS-PM apart from OLS regression. Rönkkö and Evermann also showed analytically that the relationships between composites estimated in PLS are influenced not only by the indicator weights, but also by correlations between the measurement errors of the indicators, which are likely to be nonzero in empirical research (see also Rönkkö, in press). They supplemented these analytical results with a small simulation that demonstrated the deleterious effects of correlated measurement errors on PLS-PM estimates. Comparatively, both

SEM and path analysis with summed scales provided unbiased estimates in the presence of correlated measurement errors.

Henseler et al. acknowledged that PLS-PM does not eliminate the effects of measurement error. Nonetheless, they pointed out that forming composites with multiple indicators provides some adjustment for unreliability and further argued that PLS further reduces measurement error by assigning larger weights to more reliable indicators, which they regarded as indicators that enhance the "predictive relevance" of the composite. Henseler et al. criticized the Rönkkö and Evermann simulation due to the small number of conditions it comprised and reported a simulation that contained a larger number of conditions deemed to be more representative of situations involved in empirical research. Based on this simulation, Henseler et al. concluded that composites derived using PLS mode A generally yield higher reliabilities than unweighted composites and the best single indicator used to form a composite (i.e., the indicator with the highest loading) and also produced much higher reliabilities than PLS mode B.

The apparent disagreements between Rönkkö and Evermann and Henseler et al. arose largely because they emphasized different sets of issues. With regard to the simulations, Rönkkö and Evermann focused on the effects of correlated measurement errors on reliabilities and path estimates in PLS models. In contrast, Henseler et al. addressed differences in reliabilities yielded by PLS, unweighted composites, and the best single indicator, making no mention of correlated measurement errors. Thus, the Henseler et al. simulation was not equipped to challenge the conclusions of Rönkkö and Evermann regarding the effects of correlated measurement errors, and the Rönkkö and Evermann simulation did not contradict the results of Henseler et al., given that the conditions of the Rönkkö and Evermann simulation that were included in the Henseler et al. simulation yielded essentially the same results. Thus, the conflicting views of Rönkkö and

Evermann and Henseler et al. primarily involve the conditions that should be included in their respective simulations, as opposed to the conclusions of the simulations themselves.

Turning to the results of the simulations, Henseler et al. concluded that "*PLS mode A clearly outperforms sum scores*" (emphasis in original) when indicator loadings vary widely, when the composite variables in the model are at least moderately related (i.e., $\beta = 0.5$), and when sample sizes are relatively large (i.e., 500 vs. 100), representing conditions not included in the Rönkkö and Evermann simulation. However, differences in average reliabilities yielded by these conditions were very small in absolute terms, ranging from 0.004 to 0.005. These differences are small enough to question whether the superiority of PLS mode A over summed scores is substantively important. In the remaining conditions, reliabilities were higher for summed scores than for PLS mode A, but again the differences were rather small, with values of 0.004, 0.019, and 0.145. More to the point, the results of both simulations were evaluated by subjectively comparing reliabilities across conditions, which raises questions as to whether the observed differences are meaningful. The results of the simulations would be more conclusive if confidence intervals were constructed around the reliabilities to more clearly evaluate their differences (Kelley & Cheng, 2012; Maydeu-Olivares, Coffman, García-Forero, & Gallardo-Pujol, 2010; Padilla, & Divers, 2013), and if the effects of the different reliabilities on model parameters were statistically compared (Yetkiner & Thompson, 2010).

Another limitation of the Rönkkö and Evermann simulation is that the true population model used to generate the data did not contain correlated measurement errors. Rather, the simulation was only equipped to study the effects of nonzero measurement error correlations that arise by chance due to sampling variability (see also Rönkkö, in press). The negative effects of these measurement error correlations should vanish as the sample size becomes larger, for both

PLS-PM and SEM. Thus, the results obtained by Rönkkö and Evermann might be attributed more to the small sample size (N = 100) used in their simulation rather than the relative ability of the different approaches to handle measurement errors that are correlated in the population. When correlations among measurement errors exist in the population and are ignored, larger sample sizes only magnify the ability of overidentification tests, such as the chi-square, to detect the misspecification, and parameter estimates and standard errors will likely be biased. Indeed, in most simulation studies examining the effects of correlated measurement errors, the true population model explicitly contains non-zero error correlations (e.g., Cole et al., 2007; Reddy, 1992; Saris & Aalberts, 2003; Westfall, Henning, & Howell, 2012). Thus, if Rönkkö and Evermann had adopted this approach, the results of their simulation would have been more informative.

As a further observation regarding the Henseler et al. simulation, we see little need to demonstrate that reliability is generally higher for a composite than for a single indicator. This point is well established in the psychometric literature (e.g., Nunnally, 1978; Nunnally & Bernstein, 1994), and it follows from formulas used to compute reliability estimates, such as Cronbach's alpha. For instance, when indicators are standardized, alpha can be computed as follows:

$$\alpha = \frac{k\bar{r}_{ij}}{1+(k-1)\bar{r}_{ij}} \tag{8}$$

where k is the number of items and $\bar{r}_{ij}$ is the average interitem correlation. Table 1 applies this formula with k ranging from 2 to 10 and $\bar{r}_{ij}$ ranging from 0.10 to 0.90 in increments of 0.10. To illustrate the comparison of the reliability of a composite with that of a single indicator, consider a scale with three items and an average interitem correlation of 0.50. As shown in Table 1, this scale has a reliability of 0.75. If the items had equal loadings, then each loading would equal the square root of the average interitem correlation, or $0.50^{1/2} = 0.71$, and the reliability of each item would equal the square of its loading, or $0.71^2 = 0.50$. If the items had different loadings, then an item with a loading of 0.87 would have a reliability of 0.75, the same as that of the scale, although the loadings of the remaining items would have to be lower to maintain the average interitem correlation of 0.50 (one possible pattern of loadings is 0.87, 0.62, and 0.62). Thus, the fact that composites tend to have higher reliabilities than individual items can be taken as a foregone conclusion.

We should also note that all of the reliabilities reported in the two simulations are based on the common factor model, not composite factor model (Nunnally, 1978; Nunnally & Bernstein, 1994). Elsewhere in their rebuttal, Henseler et al. critiqued Rönkkö and Evermann for relying on the common factor model, pointing out that the results of common factor models (SEM) and composite factor models (PLS-PM) cannot be directly compared, given that they estimate different underlying population parameters. Despite these admonitions, Henseler et al. invoked the common factor model to estimate reliability associated with PLS-PM. Thus, their simulation does not address how measurement error is actually represented in the composite model of PLS-PM.

Furthermore, debating whether PLS-PM or summed scales yield higher reliabilities seems rather superfluous in the common factor context, because neither of these approaches avoids the effects of measurement error. PLS-PM and summed scales both involve composites containing the measurement error carried by the indicators, which is not somehow purged when the composite is formed. Certainly, forming composites provides some relief from the effects of measurement error, given that reliability increases as the number of items in a composite increases, as shown in Table 2. However, this increase in reliability occurs at a decreasing rate, and error is never completely eliminated, regardless of the number of items or the magnitude of the average interitem correlation. Indeed, Rönkkö and Evermann conclude their discussion of reliability by noting that: “The options available for reducing the effect of measurement error with composite variables are limited because any linear composite of indicators that contain error will also be contaminated with error” (p. 436). In a similar vein, Henseler et al. acknowledge that “PLS does not completely eliminate the effects of measurement error,” and although they claim that PLS reduces these effects substantially, these benefits depend entirely on the number of indicators and their intercorrelations, as made obvious by Table 2. Thus, we conclude that Rönkkö and Evermann and Henseler et al. generally agree that PLS-PM does not eliminate the effects of measurement error, a point we think is beyond dispute. Although the amount of measurement error is a matter of degree, it necessarily hampers the ability of PLS-PM to accurately estimate the parameters of common factor models. Moreover, an imperfectly measured composite in a model necessarily propagates bias to other composites, even if perfectly measured, via the relationships among the composites, thereby undermining all structural model estimates (Antonakis et al., 2010).

Stepping back from the details of the two simulations, we believe the results of both simulations would have been more useful if they had explicitly addressed the distinctions between the composite and common factor models (Goodhue, Lewis, & Thompson, 2012a; Marcoulides & Chin, 2013). Under the common factor model, the best approach is SEM with latent variables, which is undeniably superior to PLS-PM and summed scales. This superiority arises from the fact that SEM includes parameters that segregate measurement error from the model that relates the latent factors, whereas PLS-PM currently does not have this capability. In addition, correlated measurement errors can be incorporated into SEM, thereby allowing researchers to explicitly compensate for the types of effects demonstrated in the Rönkkö and Evermann simulation (see also Rönkkö, in press). It must be stressed, however, that the addition of correlated errors to common factor models should be accompanied by an explicit theoretical and/or methodological rationale, rather than done in an uncritical manner simply to improve statistical fit (Boomsma, 2000; Cote, & Greenberg, 1990; Gerbing & Anderson, 1984; Saris & Aalberts, 2003). Although the new PLSc method improves the correspondence between SEM and PLS-PM estimates when estimating common factor models (Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press), it cannot accommodate correlated measurement errors, which frequently arise in application (Cole et al., 2007; Lee & Antonakis, 2012; Reddy, 1992; Saris & Aalberts, 2003; Westfall et al., 2012). Therefore, there seems to be little point in continuing to pit SEM and PLS-PM against each other when evaluating common factor models. Rather, it seems most prudent to just leave common factor model territory solely to SEM.

If the composite factor model is assumed (Bentler & Huang, 2014; Rigdon, 2012), then it can indeed be useful to compare PLS-PM, and possibly other component-based modeling techniques, to SEMs parameterized to incorporate composite variables (e.g., Dolan, 1996; Dolan et al., 1999; McDonald, 1996; Treiblmaier et al., 2011). In this manner, the underlying model would be the same for both methods, which would allow meaningful comparisons of their relative performance. Naturally, when comparing the approaches within a composite-based modeling scheme, unreliability and measurement error should not be evaluated and compared using procedures rooted in the common factor model (Rigdon, 2012). Although there is a current paucity of reliability indices appropriate for composite-based modeling, methods have recently been devised to adjust for irrelevant variation in composite predictors and thereby improve explanatory power in the PLS-PM context. More specifically, the *OnPLS* approach partitions the total variability into three components: (a) a global component shared among all theoretically connected blocks of observed variables, which is essentially the structural model of theoretical interest; (b) a locally joint component that represents variability shared between some but not all of the blocks; and (c) a unique component that reflects variance specific to a single block (Löfstedt, Eriksson, Wormbs, & Trygg, 2012; Löfstedt, Hoffman, & Trygg, 2013; Löfstedt, Hanafi, & Trygg, 2013; Löfstedt & Trygg, 2011). Although this approach does not completely purge measurement error from each construct, as accomplished with the common factor model, it disattenuates the estimates of the core hypothesized relationships by removing all variation that is irrelevant to prediction, which is a major advance in the component-based modeling domain. The OnPLS strategy could be further strengthened by using the IVE approaches discussed in the previous section, which can improve the consistency of estimation in the presence of measurement error (Abarin & Wang, 2012; Hardin & Carroll, 2003). In addition, combining the OnPLS method with existing SEM strategies for incorporating composites (e.g., Dolan, 1996; Dolan et al., 1999; McDonald 1996; Treiblmaier et al., 2011) could generate a powerful and versatile component-based modeling technique, which would provide the benefits of SEM's more well-developed arsenal of estimation and testing routines. Further empirical evaluation and comparison of SEM, PLS-PM and other approaches to estimate component-based path models are essential to verify these possibilities.

3. Is PLS-PM Capable of Validating Measurement Models?

Rönkkö and Evermann questioned the utility of PLS-PM for validating measurement models. Their critique focused on the criteria commonly employed by studies that examine measurement models using PLS-PM, such as the composite reliability (CR) and the average variance extracted (AVE), as well as other criteria used less frequently, such as the relative goodness-of-fit (GoF) index and the standardized root mean square residual (SRMR). Rönkkö and Evermann cited research which they claim debunks the use of the composite reliability (CR), average variance extracted (AVE; Aguirre-Urreta, Marakas, & Ellis, 2013) and GoF indices (Henseler & Sarstedt, 2013) for PLS-PM models. To bolster these claims, Rönkkö and Evermann conducted a simulation to evaluate the ability of the CR, AVE, AVE-highest squared correlation (i.e., the maximum rather than average variance extracted for a set of indicators), the relative GoF, and the SRMR to detect various types of model misspecifications. Rönkkö and Evermann found that none of the criteria they examined dependably identified measurement models that were incorrectly specified. From these results, Rönkkö and Evermann concluded that "the measurement model should never be evaluated based on the composite loadings produced by PLS or any statistic derived from these [*sic*]" (p. 438)**,** encouraging researchers to instead rely on established alternatives, such as chi-square tests of exact fit and common factor analysis (e.g., Nunally, 1978),

Henseler et al. countered by arguing that the Rönkkö and Evermann simulation was replete with errors, ranging from mistakenly equating PLS-PM with the common factor model to miscalculating the CR, AVE, and SMSR and misreporting their results. They also criticized Rönkkö and Evermann for not explicitly comparing PLS-PM and covariance-based SEM as methods for validating measurement models. Henseler et al. conducted a simulation intended to address these errors and omissions and included additional evaluation criteria, such as chi-square tests of exact fit. From the results of their simulation, Henseler et al. concurred with Rönkkö and Evermann regarding the shortcomings of the CR, AVE, and relative GoF for detecting measurement model misspecifications. In contrast, the tests of exact fit and the SRMR performed well, with somewhat better performance for covariance-based SEM than for PLS-PM. Henseler et al. discounted the apparent superiority of covariance-based SEM because it suffered from nonconvergence and improper solutions (i.e., Heywood cases), whereas PLS-PM did not. In the end, Henseler et al. recommended the chi-square test of exact fit and SRMR yielded by PLS-PM for detecting measurement model misspecification.

We cannot adjudicate the accuracy of the simulations reported by Rönkkö and Evermann and Henseler et al., because doing so would require access to the raw output of the simulations. Nevertheless, we can assess what the criteria examined in the simulations are equipped to detect and whether they should, in principle, uncover the types of model misspecifications included in the simulations. To frame this assessment, we first distinguish between two distinct properties of a measurement model: (1) the magnitudes of the estimated parameters in the model; and (2) the degree to which the model fits the data. The first property influences the CR, the AVE, and the relative GoF, which are essentially different ways of summarizing explained variation in the indicators. For example, assume a single-factor model with the variance of the factor fixed to unity. With this specification, the formula for the CR is as follows (Jöreskog, 1971):

$$CR = \frac{\left(\sum_{i=1}^{p} \lambda_i\right)^2}{\left(\sum_{i=1}^{p} \lambda_i\right)^2 + \sum_{i=1}^{p} \theta_{\delta ii}} \tag{9}$$

where p is the number of observed indicators (i = 1 though p), the λ_i are the indicator loadings, and the $\theta_{\delta ii}$ are the measurement error variances. The formula for the AVE draws from the same model parameters (Fornell & Larcker, 1981):

$$AVE = \frac{\sum_{i=1}^{p} \lambda_i^2}{\sum_{i=1}^{p} \lambda_i^2 + \sum_{i=1}^{p} \theta_{\delta ii}} \tag{9}$$

Application of these formulas for the CR and AVE in PLS-PM requires certain modifications, as outlined by Aguirre-Urreta et al. (2013). The GoF and relative GoF, which were designed specifically for PLS-PM, are functions of the item loadings and the variance explained by the structural equations in a model (Henseler & Sarstedt, 2013). Thus, all four of these measures are determined by the magnitudes of the parameter estimates for a model and are insensitive to how well the model fits the sample data.

The second property of a measurement model, which concerns model fit, is captured by the chi-square test of exact fit and the SRMR. These and other global fit statistics provide summaries of how well the model structure – that is, the number of constructs and the pattern of free and constrained parameters – reproduces the relationships among the observed variables, irrespective of the strength of those relationships (Marsh, Hau, & Grayson, 2005). To illustrate further, the conventional ML chi-square statistic in SEM is computed as $(N – 1)F_{ML,}$ where F_{ML} is the minimum value of the following discrepancy function that is used to guide the estimation of model parameters (Bollen, 1989; Hayduk, 1987):

$$F_{ML} = \log|\Sigma| - \log|S| + trace(S\Sigma^{-1}) - p\,. \tag{10}$$

In this equation, S and Σ are, respectively, the observed and model-implied covariance matrices of the observed variables, | . | denotes the matrix determinant, *trace* is an operator that sums the diagonal elements of a matrix, and p is the number of observed variables. If the model is properly specified and additional supporting assumptions are met (i.e., a large sample size and multivariate normality), then S and Σ will be equivalent (within sampling variability), the $\log|S|$ and $\log|\Sigma|$ terms will cancel each other out, and the product of $S\Sigma^{-1}$ will be an identity matrix with *trace* = p; the quantity $(N - 1)F_{ML}$ will be distributed as a central χ^2 variate on $p - q$ degrees of freedom, where q is the number of estimated parameters in the model. Therefore, the chi-square statistic provides a sharp test of whether the data conform to the structure of the hypothesized model (a comparable chi-square statistic has been developed for PLS-PM; see Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press).

Like the chi-square, the SRMR provides an overall summary of the discrepancies between S and Σ (in standardized form), as follows (Wang & Wang, 2012):

$$SRMR = ((\sum_j \sum_k r_{jk}^2) / p^*)^{1/2}, \tag{11}$$

where r_{jk} is the difference in the corresponding elements of the sample and model-implied correlation matrices, and p^* is the number of non-redundant elements in each matrix ($p^* = p(p + 1)/2$). Note, however, that the SRMR is only a descriptive goodness-of-fit index rather than a test of causal specification (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

Unfortunately, the simulations conducted by Rönkkö and Evermann and Henseler et al. did not respect the distinction between the magnitudes of model parameters and the overall model. Rathre, both simulations manipulated only the model structure, not the sizes of the parameters in the model , such that a population measurement model was estimated as specified and with various types of misspecifications (i.e., modifying the numbers of constructs and/or the pattern of free and constrained parameters). As such, it can be assumed *a priori* that the simulation results would point to the test of exact fit and the SRMR as the optimal indicators of specification error, given that these criteria are designed to detect precisely what the simulations manipulated. Measures that reflect the magnitudes of model parameters, such as the CR, AVE, GoF and relative GoF, are incapable of reflecting how well the model reproduces the data. To be sure, parameter estimates can be indirectly affected by whether a model is correctly specified, given that parameter estimates in misspecified models are often distorted (Hayduk et al., 2005; Kolenikov, 2011; McIntosh, 2007; Saris, Satorra, & van der Veld, 2009; Yuan, Marshall, & Bentler, 2003), but the direction of this distortion can be either upward or downward. Therefore, researchers should not interpret measures based on parameter magnitude as indicating model fit, which is the province of the chi-square test and SRMR. Similarly, parameter magnitude should not be taken as implying model misspecification, as a measurement model could adequately reproduce the observed covariances, yielding a non-significant chi-square statistic, but have parameters that are large or small in magnitude. This point is illustrated by a single-factor model with three indicators, which is saturated (i.e., has zero degrees of freedom) and therefore fits any covariance matrix perfectly, regardless of the magnitudes of the covariances and the associated item loadings and measurement error variances.In sum, global fit statistics and measures that reflect the magnitudes of loadings and other model parameters should always be employed in a complementary fashion when assessing measurement models, and only for their intended purposes.

With respect to using PLS-PM or covariance-based SEM to detect model misspecification, Henseler et al. argued in favor of PLS-PM due to the high proportion of nonconvergent runs and improper solutions (i.e., Heywood cases) obtained when SEM was used. However, Henseler et al. set the sample size for their simulations at 100 cases. Previous research has shown that nonconvergence and improper solutions are more common when SEM is applied to small samples, such as those analyzed by Henseler et al., due to the effects of sampling error (Dillon, Kumar, & Mulani, 1987; Fan, Thompson, & Wang, 1999; MacCallum, Widaman, Zhang, & Hong, 1999). These problems tend to decrease as sample size increases, which would likely eliminate any apparent benefit of PLS-PM over SEM. Indeed, it strikes us as odd that Henseler et al. did not vary sample size in their simulations that addressed measurement model validation, given that they emphasized sample size as "one of the most important variables in a simulation" (Paxton et al., 2001, p. 294) in their discussion of the reliability of PLS-PM composites relative to summed scores.

It is also possible to interpret Henseler et al.'s comparative results on convergence behavior and solution propriety as more strongly supporting the ability of ML-based SEM to detect model misspecifications. In the SEM context, nonconvergence of the ML estimator is viewed as a first sign of model misspecification (Boomsma & Hoogland, 2001), as are Heywood cases (Chen, Bollen, Paxton, Curran, & Kirby, 2001; Kolenikov & Bollen, 2012). To be sure, nonconvergence and improper solutions can arise even for properly specified models, but the Henseler et al. simulation clearly demonstrates the combined effects of misspecification and small sample size. In particular, the number of nonconvergent runs increased from 3.6% for the true model (Model 1, Figure 4) to 11.4% (Model 2, Figure 4), 13.6% (Model 3, Figure 4) and 16.8% (Model 4, Figure 4) when various misspecifications were introduced. Furthermore, when considering only the convergent runs, no inadmissible ML solutions were reported under the true model or Model 2, whereas Heywood cases occurred in more than 60% of the solutions obtained for Model 3; for Model 4, up to 100% depending on the particular statistical package used. In addition, the Type II error rates for the SEM vs. PLS-PM χ^2 statistics were: 93.2% vs. 99.8% (Model 2); 0.0% vs. 19.9% (Model 3); and 0.0% vs. 4.1 (Model 4). Altogether, these findings demonstrate that ML-based SEM did a better job of signaling model misspecification than PLS-PM, which "always converged" and yielded proper solutions, as well as more frequently accepted the misspecified models.

Stepping back from the particular evaluation criteria considered by Rönkkö and Evermann and Henseler et al., we have some reservations about the meaning and interpretation of measurement models in PLS-PM, particularly given the frequent and inappropriate borrowing of concepts from the common factor analysis framework. Composite-based models do not address the relationships between measures and constructs as these terms are usually conceived (Bentler, 1982; Borsboom, Mellenbergh, & Van Heerden, 2003; Edwards & Bagozzi, 2000), because the "latent" variables in PLS-PM are not latent in the sense of being unobserved, but instead are weighted composites of observed indicators. Note that this operational definition of the constructs in PLS-PM holds even when using Mode A estimation (i.e., "reflective measurement"), which is an attempt to more closely mimic the common factor model (cf. Fattore, Pelagatti, & Vittadini, 2012; Rigdon, 2012). Therefore, each indicator "loading" in a PLS-PM measurement model represents a part-whole relationship, because the indicator is part of the composite itself. More specifically, when using Mode A estimation, the loadings are akin to item-total correlations (Nunnally, 1978); under Mode B estimation (i.e., "formative measurement"), the loadings are based on the multiple regression of the composite variables on their indicators (Hanafi, 2007; Tenenhaus, Esposito, Chatelin, & Lauro, 2005; Esposito Vinzi, Trinchera, & Amato, 2010). As such, measurement models in PLS-PM essentially involve relationships between one type of manifest variable (i.e., individual indicators) and another type of manifest variable (i.e., weighted composites of indicators). Researchers interested in examining relationships between measures and unobserved constructs are better served by SEM, in which latent variables cannot be reduced to weighted composites of observed variables (Bentler, 1982). Thus, it seems misguided to view composite-based models through a common factor lens, a problem which characterizes the bulk of the methodological and applied literature on PLS-PM. In order for PLS-PM to serve the purposes of composite-based modeling, an essential requirement is “a complete and consistent approach to measurement which is factor-free” (Rigdon, 2012, p. 355),

Finally, when validating measurement models, researchers should consider evidence that goes beyond the general assessment criteria considered by Rönkkö and Evermann and Henseler et al. For example, in both PLS-PM and SEM applications, global fit tests should always be accompanied by local tests (Bera & Bilias, 2001; Grace et al., 2012; Saris, Satorra, & van der Veld, 2009; Yuan, Kouros, & Kelley, 2008) and additional diagnostics such as fitted residuals (i.e., the differences between the elements of the sample and model-reproduced covariance matrices). Other essential features of measurement models include the validity of individual items and the convergent and discriminant validity of the factors that constitute measurement models. Under the common factor model, this information can be obtained by thoroughly examining item loadings, measurement error variances, and factor correlations, which should all be routinely considered in studies intended to validate measures (Jackson, Gillaspy, & Purc-Stephenson, 2009). Further recommendations for evaluating common factor models are available elsewhere (Bagozzi & Phillips, 1991; Bollen, 1989; Brown, 2006; Harrington, 2009; Klineb, 2013; Raykov & Marcoulides, 2010; Schumacker & Lomax, 2010). For composite factor models, corresponding guidelines are in short supply due to the fundamental differences between the two measurement frameworks. In fact, it has been suggested that conventional notions of validity might not even be relevant for composite constructs, because composite indicators need not always have “conceptual unity” (Bollen, 2011, p. 372; see also Bollen & Bauldry, 2011). As a recourse, PLS-PM practitioners could recast the assessment of measurement quality in terms of the model’s predictive ability, concentrating on examining the predictive utility of the composite variables (cf. Rigdon, 2012). However, the interpretation of the composites themselves would remain problematic, particularly when they combine indicators that are conceptually heterogeneous (Edwards, 2011; Hardin & Chang, 2013; Hattie, 1985; Howell, Breivik, & Wilcox, 2013).

Shifting focus to prediction comes with its own challenges. First, the specification of the model (i.e., the number of constructs and the pattern of free and fixed parameters) would still need to be verified through global and local fit tests before any predictions can be trusted (Antonakis et al., 2010; Grace et al., 2012; Hayduk et al., 2005, 2007; McIntosh, 2007). Second, composite-based models will be particularly susceptible to overfitting (Hawkins, 2004; Hegdé, 2010; Marewski & Olsson, 2007). Because the estimates of structural parameters in these models capture both the theoretical process of interest and sampling variability in the particular data used for analysis, the model will generally achieve better within-sample prediction (i.e., predictive ability within the data set used to estimate the parameters) than out-of-sample prediction (i.e., predictive ability in data not used in estimation). Given that the latter is the most important indicator of a models’ predictive utility (Meese & Rogoff, 1983), cross-validation should be used to assess the extent to which modeling results generalize to independent data sets (cf. Chin, 2010; Arlot & Celisse, 2010). In addition, the OnPLS (Löfstedt et al., 2013) and IVE methods (Abarin & Wang, 2012; Hardin & Carroll, 2003) can be used to help compensate for the impacts of measurement error on overfitting. Such approaches are particularly important given that high-dimensional, low sample size data sets often used to justify PLS-PM are particularly prone to overfitting and accompanying Type I error inflation (Fan, Guo, & Hao, 2012; Forstmeier & Schielzeth, 2011; Subramanian & Simon, 2013).

4. Does PLS-PM Provide Valid Inference on Path Coefficients?

Rönkkö and Evermann critiqued the use of bootstrapped tests of significance for PLS-PM path coefficients on two grounds: (1) the critical ratios (i.e., parameter estimates divided by their bootstrap standard errors) are referred to the *t* distribution even when parameter estimates are not normally distributed; and (2) the bootstrap distribution of the parameter estimates deviates from the corresponding analytical sampling distribution. Rönkkö and Evermann conducted a simulation revealing that the accuracy of inferences in PLS-PM may be compromised in certain situations. Using a two-construct population model with no structural relationship (i.e., $\beta = 0.0$), Rönkkö and Evermann found a bimodal (i.e., two-peaked) distribution for the path coefficient at a sample size of 100. Moreover, the bootstrapped distribution of the PLS-PM path coefficient was markedly different from the original sampling distribution obtained from the Monte Carlo replications. These demonstrations suggested that PLS tests of inference can be biased and inconsistent.

Henseler et al. responded by replicating and extending the simulations conducted by Rönkkö and Evermann. As before, the sampling distributions of path coefficients of both $\beta = 0.0$ and $\beta = 0.3$ in the same two-construct path model were bimodal at a sample size of 100. However, modifying certain aspects of the simulation design used by Rönkkö and Evermann produced sampling distributions that were unimodal. Specifically, when holding the effect size constant at $\beta = 0.3$, unimodality was achieved by either: (a) increasing the sample size to 500; (b) making the loadings more heterogeneous; or (c) adding two constructs to the model. In addition, altering the original simulation by simply increasing the effect size to $\beta = 0.5$ also yielded a unimodal distribution. Furthermore, Henseler et al. found that three different strategies for constructing bootstrap-based confidence intervals (i.e., normal bootstrap, percentile confidence intervals, and bias-corrected and accelerated confidence intervals) provided impressive overall Type I and Type II error control, even at $N = 100$.

We agree with Henseler et al. that coefficient distributions for PLS-PM are not necessarily bimodal. However, because parameter distributions can deviate from normality under certain conditions, the routine use of the *t* distribution to test parameters is questionable. Furthermore, the Henseler et al. simulation did not disprove Rönkkö and Evermann's core finding, which is that parameter distributions are bimodal under the null hypothesis, such that $\beta = 0.0$. Under this condition, the Henseler et al. simulation replicated the results of Rönkkö and Evermann. The remaining conditions examined by Henseler et al. specified nonzero effect sizes in the population, which consistently generated parameter distributions that were unimodal. These results suggest that PLS-PM inference might be well served by adopting an *alternative hypothesis significance testing* (AHST) framework, in which bootstrap resampling would be performed with respect to values other than the conventional null hypothesis (cf. Rodgers & Beasley, 2013). However, in cases where the researcher lacks the requisite theory or prior empirical evidence to postulate nonnull hypotheses, it may be most prudent to invoke the null. Therefore, additional work is needed to more fully determine the behaviour of PLS-PM path coefficients under the null in order to delineate the specific conditions for which inference using the usual null hypothesis-based *t*-distribution can be justified.

If the researcher does not wish to presume normality in a given application, non-parametric methods of hypothesis testing are available that make no distributional assumptions. For instance, permutation (or randomization) tests have already been introduced into PLS-PM to compare parameter estimates across multiple groups (e.g., Chin & Dibbern, 2010; Crisci & D'Ambra, 2012). This approach is a promising alternative to the usual bootstrap procedure. Nonetheless, the bootstrapped confidence intervals demonstrated by Henseler et al. appear reasonably robust to violations of normality and divergence between analytical and bootstrap sampling distributions, at least for the relatively simple models examined by Henseler et al. Future simulation work is needed that considers more complex models, different sample sizes, and additional violations of assumptions to determine when this approach might break down.

Moreover, significance tests are only meaningful if the estimates can be trusted to be consistent, regardless of the robustness of the inference procedure (Freedman, 2006; King & Roberts, 2012). As discussed earlier, the consistency of parameter estimates depends on whether a model is correctly specified, as evidenced by passing global and local tests of parameter constraints, as well as adjusting for endogeneity of predictors. Clearly, bootstrapping an inconsistent estimate will not make inferential tests consistent.

Beyond these issues, it is important to point out another common modeling situation in which bootstrapped (and permutation-based) inference in PLS-PM is likely to deteriorate, and for which no remedy has yet been devised in the PLS-PM context: analysis of data from a complex survey designs. Complex survey designs are used to collect data in situations where simple random sampling, which is the ideal for conventional statistical inference, is not feasible for practical or ethical reasons (de Leeuw, Hox, & Dillman, 2008; Heeringa, West, & Berglund, 2010; Lehtonen & Pahkinen, 2004). Complex survey designs typically have two main features: stratification and clustering. Stratification involves partitioning the sample into independent groups and then sampling within each group. This strategy helps ensure that groups on which information is desired are adequately sampled. For instance, if a survey focuses on the attitudes and perceptions of managers at different levels of an organization, managers could be grouped according to level prior to drawing the sample. Whereas this procedure renders the sampling process more efficient than simple random sampling, not every member of the population has an equal chance of being sampled, thereby violating the assumption that observations are identically distributed and leading to biased parameter estimates (Raghav & Barreto, 2011). In contrast to stratification, clustering refers to naturally-occurring groupings of lower-order sampling units within higher-order units (e.g., employees within organizations, students within schools, patients within hospitals). Given that responses within clusters tend to be intercorrelated, the assumption that observations are independent is compromised, resulting in underestimation of standard errors and inflated Type I error rates (Barreto & Raghav, 2013; McCoach & Adelson, 2010).

The problems arising from stratification and clustering can be addressed in various ways (Hahs-Vaughn, McWayne, Bulotsky-Shearer, Wen, & Faria, 2011; Lumley, 2011; Osbourne, 2011). For instance, observations can be weighted according to their selection probabilities to obtain accurate parameter estimates (Pfeffermann, 1993, 1996). The calculation of sampling weights need not fall on the analyst, because these weights are typically computed by survey methodologists and supplied with the data set. Furthermore, appropriate standard errors can be derived using resampling techniques, such as the bootstrap or jackknife (Kolenikov, 2010), or by applying post-estimation corrections (Cameron, Gelbach, & Miller, 2011; Hedges, 2007, 2009; Thompson, 2011). Alternatively, researchers can use statistical modeling techniques that directly incorporate features of the survey design as components of the analysis (i.e., strata, clusters), in order to ensure that estimation and inference are not compromised. For example, *multilevel modeling* (MLM) allows coefficients to potentially vary randomly across the clusters (Goldstein, 2011; Hox, 2010; Snijders & Bosker, 2012) and can also compute standard errors that are robust to common violations of modeling assumptions (e.g., normality of level 2 errors and homogeneity of variance across clusters; Hox, 2010; Maas & Hox, 2004a, 2004b). Combining MLM with bootstrap resampling has been shown to further improve the accuracy of inference (Kovacevic, Rong, & You, 2006; Pierre & Saidi, 2008; Roberts & Fan, 2004; Seco, García, García, & Rojas, 2013; van der Leeden, Meijer, & Busing, 2008).

Implementing these methods would address some of the shortcomings of current applications of PLS-PM. The fact that resampling is already the method of choice for PLS inference should facilitate the use of the adjustments described here. However, the resampling procedures available in PLS-PM software assume that the data are collected via simple random sampling, such that observations are both independent and identically distributed (e.g., Kock, 2013; Ringle, Wende, & Will, 2005; Sanchez & Trinchera, 2013). When complex survey data is involved, the resampling strategy must mimic the original sampling scheme (Aidara, 2013; Antal & Tillé, 2011; Pal, 2009; Preston, 2009), and PLS resampling routines should be modified accordingly. Similarly, the new permutation approach to PLS-PM inference (cf. Chin & Dibbern, 2010; Crisci & D'Ambra, 2012) would also need to be adjusted for complex survey data (Pesarin & Salmaso, 2010). Guidance for incorporating MLM and design-based resampling techniques into PLS-PM can be found in the SEM field, which has made considerable advances in both of these areas (e.g., Bai & Poon, 2010; Hox, 2013; Kaplan, Kim, & Kim, 2009; Oberski, 2013; Rabe-Hesketh, Skrondal, & Zheng, 2007; Stapleton, 2008; Wu & Kwok, 2012).

5. Is PLS-PM Advantageous at Small Sample Sizes?

Rönkkö and Evermann reviewed and evaluated previous studies addressing the performance of PLS-PM with small sample sizes on various criteria (e.g., convergence, bias, efficiency, and power). They argued that the apparent advantages of PLS-PM with small sample sizes can be attributed to: (1) ignoring the effects of chance correlations among measurement errors, which inflates parameter estimates; and (2) the use of a *t* distribution for NHST when the coefficient distributions are not normal, which leads to increased Type I error rates. To bolster this argument, Rönkkö and Evermann reported a small simulation that compared the performance of PLS-PM, SEM, and path analysis with summed scales across sample sizes of 25, 50, and 100. They found that, when the population effect size was zero, PLS-PM produced bimodal parameter estimates with modes that were positive and negative, whereas SEM and the summed scale approach yielded estimates centered at zero. When the population effect was positive, PLS-PM generated estimates that progressively exceeded the population effect as sample size decreased, whereas SEM produced estimates centered at the population value, and summed scales yielded estimates slightly below the population value. Based on these results and related empirical evidence from the literature (Chin & Newstead, 1999), Rönkkö and Evermann concluded that PLS-PM has no advantages when applied to small samples and recommend that researchers should maximize sample size. If this is not possible, they suggested making use of specialized small sample estimation techniques developed in the SEM context or parceling indicators to reduce the overall number of observed variables in the model.

Henseler et al. agreed with Rönkkö and Evermann's observation that many researchers apparently assume that PLS-PM is amenable to small sample sizes. They then compared PLS-PM to SEM with regard to both statistical power and convergence. They cited evidence from the literature indicating that, in small samples, SEM tends to produce smaller standard errors than those yielded by PLS-PM and argued that any conflicting findings in this regard are statistical artifacts stemming from fundamental differences between composite and common factor model parameters. Furthermore, Henseler et al. reiterated their earlier simulation findings on the high frequency of nonconvergence and improper solutions in SEM under misspecified measurement models and a sample size of 100. From these observations, Henseler et al. concluded that, with small samples, PLS-PM can be successfully applied when other methods fail, referring particularly to SEM.

The different conclusions drawn by Rönkkö and Evermann and Henseler et al. seem to result from the different weights they place on the criteria typically used to evaluate the performance of analytical techniques with small samples. Foremost among these criteria are convergence, bias, efficiency, and statistical power. Certainly, an analytical approach that frequently fails to converge to a solution is problematic. However, as we previously discussed, Henseler et al.'s particular results on the small-sample convergence behavior of ML-based SEM actually demonstrated its superiority over PLS-PM in signaling misspecified measurement models. Furthermore, there are various ways to address convergence issues in SEM, such as choosing better starting values, increasing the number of iterations, and modifying convergence criteria (defaults for these factors differ across packages and need not be accepted without question). For example, the Stata program has recently introduced an advanced SEM module which has various maximization options to help increase convergence rates when using ML (StataCorp, 2013). In addition, the M*plus* package for SEM offers sophisticated optimization routines for ML such as numerical and Monte Carlo integration (Muthén & Muthén, 1998-2013), which facilitate the estimation of complex models. Future comparisons between PLS-PM and SEM should consider these newly developed algorithms, as past reliance on software defaults may have painted an overly negative picture of ML's small sample convergence behaviour with properly specified models. Although one could argue that the various convergence enhancements might potentially reduce ML's first line of defense against misspecified models, the judicious use of global and local fit tests should help offset this problem.

Regarding the effects of sample size on bias, efficiency, and power, existing simulation studies have not yielded consistent results. Some studies suggest few appreciable differences between SEM and PLS-PM (e.g., Goodhue et al., 2012), whereas other studies indicate that SEM outperforms PLS-PM (Chumney, 2013; Hulland, Ryan, & Rayner, 2010). Further evidence suggests that the small sample performance of both SEM and PLS-PMdepends on which specific aspects of the model are considered. For example, some findings indicate that PLS-PM surpasses SEM at recovering estimates and standard errors within the structural (inner) portion of the model, whereas ML performs better in this regard within the measurement (outer) aspect of the model (Sharma & Kim, 2013; Vilares, Almeida, & Coelho, 2010; Vilares & Coelho, 2013). We agree with Henseler et al. that much of the relevant literature comparing the finite sample behaviour of PLS-PM and SEM is inconclusive due to naïve comparisons of composite and common factor models (see also Marcoulides & Chin, 2013; Marcoulides, Chin, & Saunders, 2012; Rai, Goodhue, Henseler, & Thompson, 2013). Treiblmaier et al. (2011) show how to properly parameterize SEM approaches to allow unconfounded head-to-head comparisons with PLS-PM in the composite-based structural model case.

Another major limitation of existing studies comparing PLS-PM and SEM is that they have not considered advances in SEM estimation and testing procedures designed to compensate for small sample size. As Rönkkö and Evermann note, this is an active area of research. For example, several Bartlett-type (1954) corrections to the conventional ML chi-square statistic have been developed to control Type I error rates when the sample size is low relative to the number of parameters in the model (Herzog, & Boomsma, 2009; Herzog, Boomsma, &

Reinecke, 2007). Among these procedures, the Swain (1975) correction has been shown to perform particularly well (Antonakis & Bastardoz, 2013; Bastardoz & Antonakis, 2013; Herzog & Boomsma, 2009; Herzog et al., 2007; Jackson, Voth, & Frey, 2013; Wolf, Harrington, Clark, & Miller, 2013). In addition, ridge-type corrections for SEM yield consistent estimates, accurate tests of model fit, and high convergence rates in small samples (Bentler & Yuan, 2011; Jung, 2013; Yuan & Chan, 2008; Yuan, Wu, & Bentler, 2011). Futhermore, robust approaches have been developed for dealing with other suboptimal data conditions that tend to exacerbate the impact of small sample size on SEM, such as missing data (Enders, 2006, 2011; Raykov, 2012; Savalei, 2010; Savalei & Yuan, 2009; Yuan & Zhang, 2012) and non-normality (Lei & Wu, 2012; Savalei, 2010, in press; Savalei & Falk, in press). Despite the availability of these advancements, the standard practice in comparisons of PLS-PM and SEM is to implement the latter using classical ML estimation, which does not reflect SEM's current capabilities. Therefore, future comparative studies on PLS-PM and SEM should be expanded to include the above-mentioned innovations to obtain more current and conclusive results about small sample performance.

One of Rönkkö and Evermann's suggestions for handling small sample size in SEM that warrants further comment involves the use of parcels. Briefly, parceling involves reducing the size of a model by aggregating (i.e., summing or averaging) subsets of observed variables and then using the aggregates as indicators. Research on the advantages and disadvantages of parceling is ongoing (e.g., Little, Rhemtulla, Gibson, & Schoemann, 2013; Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013; Rocha & Chelladurai, 2012). One of the primary concerns with parceling is that collapsing indicators into aggregates can conceal misspecification in the measurement portion of the model, leading to overly optimistic fit statistics and inflated estimates of structural parameters. Even in the ideal case where the scales are undimensional (i.e., no correlated measurement errors or cross-loadings), the model is correctly specified, and the observed variables are multivariate normally distributed, SEM fit statistics and parameter estimates have been shown to be vary depending on how indicators are allocated to parcels (Sterba, 2011; Sterba & MacCallum, 2010). Therefore, parceling should used cautiously, and researchers should report the variability of fit statistics and parameter estimates across different allocations of indicators to parcels. Software modules are available to facilitate the construction and display of parcel-allocation distributions (Sterba, 2011; Sterba & MacCallum, 2010).

6. Can PLS-PM be Used for Exploratory Modeling?

Rönkkö and Evermann pointed out that, despite the frequent descriptions of PLS-PM as an exploratory approach to model building, most applications are actually just as confirmatory as SEM studies, such that researchers use theory to specify both the constructs and system of causal pathways *a priori*, followed by parameter estimation and model evaluation. Rönkkö and Evermann claimed that the seminal work on PLS-PM did not emphasize its exploratory potential, deeming it unsuitable for both model discovery (i.e., learning a model in a data-driven fashion when the lack of a guiding theory prevents a complete *a priori* specification of relationships) and model modification (i.e., revising an initially postulated model *a posteriori* in order to achieve better representation of the sample data). Regarding model discovery, Rönkkö and Evermann asserted that PLS-PM cannot extract patterns from data because the model must be fully specified for analysis, with each indicator assigned to one and only composite variable in the outer aspect of the model and the relations among the composites posited for the inner aspect of the model. Concerning model modification, Rönkkö and Evermann voiced concerns over PLS-PM because it does not provide overidentification tests and modification indices to detect incorrect parameter restrictions. They also pointed out that both model discovery and model modification techniques already exist in SEM, and if an hypothesized model is in doubt, limited information estimators for SEM (e.g., 2SLS) are available that are less prone to propagating the impact of specification errors throughout the model.

Henseler et al. responded by criticizing Rönkkö and Evermann's representation of the early literature on PLS-PM, citing quotes from Wold that explicitly promoted an exploratory paradigm for the method. However, Henseler et al. noted that very few researchers (about 14%) who use PLS-PM have adopted an exploratory perspective. Nevertheless, Henseler et al. asserted that applications of SEM – ubiquitously touted as a confirmatory technique – also frequently involve a substantial dose of exploration to improve model fit. Further, they claimed that PLS-PM still enables exploratory modeling because: (1) the researcher can always start with a saturated inner model (i.e., all possible composite-level paths are included *a priori*) and then remove any non-significant relationships *a posteriori*); (2) chi-square tests and other fit indices (e.g., SRMR) can be used to assess whether the model is underparameterized; and (3) cases where the common factor model does not hold support the use of PLS-PM to explore whether the composite factor model is more appropriate. Henseler et al. also countered Rönkkö and Evermann's support of SEM-based modification indices and limited information estimators (e.g., 2SLS) by noting that the former have proven unreliable, and that PLS-PM is similar to limited information estimators in terms of dampening the impact of specification errors, given that the regression equations in PLS-PM are estimated separately.

Our reflections on this exchange concentrate on the fundamental methodological issue raised, i.e., the current and potential capabilities of PLS-PM for exploratory model building. First, as to modifying an hypothesized model that does not adequately fit to the data, we concur with Henseler et al. that many SEM applications end up being partially exploratory, and that conventional SEM modification indices (MIs) show suboptimal performance. More specifically, MIs: (1) assume that the rest of the model is fully correct when considering the tenability of freeing a specific restriction (i.e., estimating the expected change in both the global chi-square statistic and parameter estimates); and (2) are prone to capitalization on chance if corrective procedures are not used (Green, Thompson, & Babyak, 1998; Green, Thompson, & Poirer, 1999, 2001; Hancock, 1999). Therefore, post-hoc model modification can actually increase rather than reduce specification errors (Fan, 2010). However, local specification checks (e.g., tests of vanishing partial correlations, IVE) can help overcome some of these limitations in both SEM and PLS-PM applications, as they evaluate each constraint independently without assuming the remaining constraints are correct (Shipley, 2000, 2003, Bollen et al., 2009).

Concerning Henseler et al.'s claim that PLS-PM is an exploratory alternative to SEM when the common factor model is untenable, we maintain that researchers need to make an *a priori* choice between the common factor and composite factor model. This decision should be based on careful consideration of whether the observed variables are *reflective* or *formative* indicators of the theoretical constructs (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000; MacKenzie, Podsakoff, & Jarvis, 2005). If the theoretical constructs can be viewed as common underlying causes of their respective indicators, then a reflective (i.e., common factor) measurement model is the obvious choice. Reflective measurement models are particularly well-suited to multi-item instruments that assess unobservable psychological constructs (e.g., anxiety, self-esteem, affect, quality of life, etc.) (Fayers & Hand, 1997, 2002; Nunnally & Bernstein, 1994; Raykov & Marcoulides, 2010). In cases where the indicators do not reflect the theoretical constructs but rather combine to produce them, a formative (i.e., composite) measurement model is appropriate (Diamantopoulos & Temme, 2013; Kline, 2013a). For instance, socio-economic status is typically conceived of as a formative construct, generated by a weighted linear combination of indicators such as income, educational attainment, occupational prestige, and neighborhood (Bollen & Lennox, 1991; National Center for Education Statistics, 2012). Formative models can be implemented using PLS-PM, other varieties of composite-based path modeling (Hwang & Takane, 2004; Hwang, 2008, 2009; Tenenhaus, 2013; Tenenhaus & Tenenhaus, 2011), reparameterizations of SEM for handling composites (Dolan, 1996; Dolan et al., 1999; McDonald, 1996), or modified versions of reflective factor models (e.g., Edwards, 2011; Treiblmaier et al., 2011). The latter two options are the most advantageous, as they allow the use of: (a) established methods of parameter estimation and fit assessment; and (b) a combination of common factors and composites within the same path model, if required.

We should add that applying limited information estimators, as recommended by both Rönkkö and Evermann and Henseler et al., could lull practitioners into a false sense of security regarding the usefulness of their models, particularly when the model is grossly misspecified. In such cases, there is little point in interpreting parameter estimates, particularly when the researcher does not know the location and magnitude of the specification errors. Again, we emphasize that local tests should be used to help ferret out the specific sources of misspecification in a model.

When a researcher has no guiding theory, such that the goal is model discovery, we concur with Rönkkö and Evermann that exploratory methods with strong analytical foundations are more widely available in SEM than in PLS-PM. The particular brand of exploratory SEM (Asparouhov & Muthén, 2009) referenced by Rönkkö and Evermann is rather limited, however, as it focuses specifically on adding cross-loadings into CFA and SEM measurement models.

Beyond this procedure, there are numerous automated search algorithms that can determine the optimal number of latent variables and system of relations between them (e.g., Landsheer, 2010; Marcoulides & Ing, 2012; Shimizu et al., 2011; Spirtes, Glymour, Scheines, & Tillman, 2010; Tu & Xu, 2011; Xu, 2010, 2012; Zheng & Pavlou, 2010). These techniques will almost invariably return several models that provide a good fit to the data. Subject matter expertise can then be used to help select the most plausible of the discovered models, which can be subjected to cross-validation using independent data. Although certainly not guaranteed to reveal the “truth” about the causal data-generating process in any given application, these exploratory SEM procedures can potentially discover meaningful and useful models that might not have been conceived in advance.

Concerning model discovery strategies for PLS-PM, we disagree with Henseler et al.’s suggestion to begin with a fully saturated inner model and then delete non-significant paths. This approach is suboptimal, as it fails to recognize that the correct model might not result by simply restricting non-significant pathways in the estimated model. Rather, the true model could differ markedly from the estimated model in terms of the number of composite variables, the pattern of free and fixed parameters, and the causal flow of the model. Therefore, when PLS-PM practitioners do not have a guiding theory, a more rigorous and sophisticated approach is needed to reveal plausible model structures

Presently, we are aware of only one technique for model discovery in the PLS-PM context, namely *universal structure modeling* (USM; Buckler & Hennig-Thurau, 2008; Turkyilmaz, Oztekin, Zaim, & Demirel, 2013), which is implemented in the *Neusrel* software package (http://www.neusrel.com/welcome/). Briefly, USM proceeds in two steps: (1) the use of PLS-based exploratory algorithms to assign observed variables to a user-specified number of composites; and (2) application of neural networks to discover the optimal system of linear, non-linear, and interactive pathways among the composites. Unfortunately, there is a dearth of published empirical research comparing USM to extant SEM and PLS-PM exploratory modeling procedures. Further work should be aimed at filling this gap, as well drawing from the vast data mining literature to construct automated model search algorithms suitable for PLS-PM (Gaber, 2010; Chakrabarti et al., 2009; Kargupta, Han, Yu, Motwani, & Kuma, 2009; Lin, Xie, Wasilewska, & Liau, 2008; Ratner, 2012). Indeed, PLS regression is often used for data mining (Allen, Peterson, Vannucci, & Maletić-Savatić, 2013; Vidaurre, van Gerven, Bielza, Larrañaga, & Heskes, 2013; Wold, Eriksson, & Kettaneh, 2010), and potential extensions of this approach should be examined. Given that sharply-defined, *a priori* conceptual frameworks are often lacking in applied research, such as when data are collected to simply describe populations or meet program reporting requirements (Boslaugh, 2007; Brady, Grand, & Powell, 2001; Trzesniewski, Donnellan, & Lucas, 2011; Vartanian, 2011), it is crucial that the methodological toolboxes of both PLS-PM and SEM practitioners contain viable approaches for exploratory modeling.

Conclusion

The present commentary has taken stock of the recent exchange between Rönkkö and Evermann and Henseler et al. on the properties and capabilities of PLS-PM. For each point of the exchange, we have summarized and critically evaluated the core arguments in light of the broader methodological and statistical literature (e.g., psychometrics, econometrics, SEM, and causal analysis). At the same time, we also offered specific recommendations for improving the ability of PLS-PM to estimate and test theoretical models. Many of these recommendations extend to other approaches for path modeling with composite variables (Hwang & Takane, 2004; Hwang, 2008, 2009; Tenenhaus, 2013; Tenenhaus & Tenenhaus, 2011), which have limitations similar to those of PLS-PM in terms of their ability to validate causal structures. Therefore, it is our hope that PLS-PM specialists and other methodologists will take the initiative to further develop and examine the viability of our ideas in future theoretical and empirical work, which may help to ultimately resolve the growing impasse between proponents and critics of PLS-PM. It should become increasingly feasible to implement and test our suggestions, given the ongoing development of open source software for custom statistical programming (e.g., R Core Team, 2013), including modules for conducting PLS-PM (Monecke, 2013; Rönkkö, 2013; Sanchez & Trinchera, 2013) and similar composite-based modeling approaches (Tenenhaus, 2013).

Looking ahead, we maintain that PLS-PM developers and practitioners should take heed of two key considerations that arose from our coverage of the two target articles and other relevant literature. First, we believe that much of the controversy surrounding the viability of PLS-PM as a statistical method can be attributed to its original development and ongoing application as a technique that attempts to imitate common factor-based SEM. As Rigdon (2012) aptly points out, “Both the method’s originators and its critics have tended to evaluate PLS path modeling in terms of what it is not” (p. 342). Moreover, the purported advantages of PLS-PM relative to SEM (e.g., reduced computational demands and superior convergence behavior, robustness to small sample size, tolerance of badly behaved distributions, exploratory capabilities in the absence of theory, etc.) also exist in the SEM domain, owing to recent theoretical and technical innovations (see Hoyle, 2012; Kaplan, 2009; Kline, 2013b; Skrondal & Rabe-Hesketh, 2004). Even the new PLSc technique does not seem capable of completely matching SEM’s estimation and testing capabilities for common factor-based modeling (e.g., Dijkstra, 2010, 2014; Dijkstra & Henseler, 2012, 2013; Dijkstra & Schermelleh-Engel, in press), because it is not equipped to deal with the ubiquitous problem of correlated measurement errors (e.g., Cole et al., 2007; Reddy, 1992; Rönkkö, in press; Saris & Aalberts, 2003; Westfall et al., 2012). Therefore, we fully support Rigdon's (2012) recommendation for PLS-PM to divorce itself completely from the factor-analytic tradition and concentrate on developing itself further as a purely composite-based statistical methodology. We also contend that this paradigm shift should be respected when using other strategies for composite-based structural modeling (Hwang & Takane, 2004; Hwang, 2008, 2009; Tenenhaus, 2013; Tenenhaus & Tenenhaus, 2011). Anything else would appear to invite stalemate and stagnation, outcomes that we would certainly like to see avoided.

Second, even within a solely composite-focused modeling league, PLS-PM and related methods will still face tough competition from SEM, which can be reparameterized in different ways to handle composite variables (Bollen, 2011, Bollen & Bauldry, 2011; Bollen & Davis, 2009; Dolan, 1996; Dolan et al., 1999; McDonald, 1996; Treiblmaier et al., 2011). Given the versatile and powerful array of model estimation and testing routines available in SEM, we are frankly skeptical at this time that either PLS-PM or any of its sister techniques could be shown to be superior in the composite case. If not, the recent calls for discontinuing the use of PLS-PM might be justified (Antonakis et al., 2010; Rönkkö, in press; Ronkko & Everman, 2013; Rönkkö & Ylitalo, 2010. However, any firm judgments in this regard should rest on a comprehensive, rigorous program of comparative research incorporating our recommendations for improving PLS-PM and similar approaches to path modeling with composite variables.

References

Abarin, T., & Wang, L. (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics*, *64*(3), 475-493. doi: 10.1007/s10463-010-0319-0

Abdi, H., Chin, W.W., Esposito Vinzi, V., Russolillo, G., & Trinchera, L. (2013). (Eds.) *New Perspectives in Partial Least Squares and Related Methods* (Series: Springer Proceedings in Mathematics & Statistics, Vol. 56). New York, NY: Springer Science+Business Media. doi: 10.1007/978-1-4614-8283-3

Abrevaya, J., Hausman, J.A., & Khan, S. (2010). Testing for Causal Effects in a Generalized Regression Model With Endogenous Regressors. *Econometrica*, *78*(6), 2043–2061. doi: 10.3982/ECTA7133

Addinsoft. (2013). XLSTAT – Statistics package for Excel. URL http://www.xlstat.com/en/

Aguirre-Urreta, M. I., Marakas, G. M., & Ellis, M. E. 2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *ACM SIGMIS Database*, *44*(4), 11-43. doi: 10.1145/2544415.2544417

Aidara, C.A.T. (2013). Bootstrap variance estimation for complex survey data: a Quasi Monte Carlo approach. *Sankhya B*, *75*(1), 29-41. doi: 10.1007/s13571-013-0062-7

Allen, G.I., Peterson, C., Vannucci, M., & Maletić-Savatić, M. (2013). Regularized partial least squares with an application to NMR spectroscopy. *Statistical Analysis and Data Mining*, *6*(4), 302–314. doi: 10.1002/sam.11169

Angrist, J.D, & Krueger, A.B. (2001). Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69–85. doi: 10.1257/jep.15.4.69

Antal, E., & Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, *106*(494), 534-543. doi: 10.1198/jasa.2011.tm09767

Antonakis J., & Bastardoz, N. (2013). *Swain: Stata module to correct the SEM chi-square overidentification test in small sample sizes or complex models*. Statistical Software Components S457617, Boston College Department of Economics.
URL http://econpapers.repec.org/software/bocbocode/s457617.htm

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C.A. (2014). What makes articles highly cited? *The Leadership Quarterly*, *25*(1), 152-179.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: a review and recommendations. *The Leadership Quarterly*, *21*(6), 1086-1120. doi: 10.1016/j.leaqua.2010.10.010

Antonakis J., Bendahan S., Jacquart P. & Lalive R. (2014). Causality and endogeneity: Problems and solutions. In D. Day (Ed.), *The Oxford Handbook of Leadership and Organizations (pp. 93-117). New York: Oxford University Press.*
Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. doi: 10.1214/09-SS054

Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*(3), 397-438. doi: 10.1080/10705510903008204

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*, 421-458.

Bai, Y., & Poon, W.-Y. (2009). Using Mx to analyze cross-level effects in two-level structural equation models. Structural Equation Modeling, 16(1), 163-178.
doi: 10.1080/10705510802561527

Barreto, H., & Raghav, M. (2013). *Understanding and Teaching Within-Cluster Correlation in Complex Surveys*. Working Paper 2013-02. Greencastle, IN: DePauw University, Department of Economics and Management.
URL http://ideas.repec.org/p/dew/wpaper/2013-02.html

Bartlett, M. S. (1954). A note on the multiplying factors for various chi-square approximations. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *16*(2), 296-298.

Bastardoz, N. & Antonakis, J. (2013). *How small can a sample size be for a structural equation model*? Congress of the Swiss Society of Psychology: Crossing Borders. Basel, Switzerland.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, *3*(1), 1-31.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, *7*(4), 465-506.

Beasley, T.M. (2008). Seemingly Unrelated Regression (SUR) Models as a Solution to Path Analytic Models with Correlated Errors. *Multiple Linear Regression Viewpoints*, *34*(1), 1-7.
URL http://mlrv.ua.edu/2008/vol34_1/Beasley-SUR.pdf

Belsley, D.A. (1992). Paring 3SLS calculations down to manageable proportions. *Computer Science in Economics & Management*, *5*(3), 157-169. doi: 10.1007/BF00426758

Bentler, P. M. (1982). Linear systems with multiple levels and types of latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems Under Indirect Observation* (pp. 101–130). Amsterdam: North-Holland.

Bentler, P. M., & Huang, W. (2014). On components, latent variables, PLS and simple methods: reactions to Ridgon's rethinking of PLS. *Long Range Planning*, forthcoming.

Bera, A. K., & Bilias, Y. (2001). Rao's score, Neyman's C (α) and Silvey's LM tests: an essay on historical developments and some new results. Journal of Statistical Planning and Inference, 97(1), 9-44. doi:10.1016/s0378-3758(00)00343-8

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

Bollen, K.A. (1996). An Alternative Two Stage Least Squares (2SLS) Estimator for Latent Variable Equations. *Psychometrika*, *61*(1), 109-121. doi: 10.1007/BF02296961

Bollen, K.A. (2011). Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models. *MIS Quarterly*, *35*(2), 359-372.

Bollen, K.A. (2012). Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology*, *38*, 37-72.
doi: 10.1146/annurev-soc-081309-150141

Bollen, K.A., & Bauer, D.J. (2004). Automating the selection of model-implied instrumental variables. *Sociological Methods & Research*, *32*(4), 425-452.
doi: 10.1177/0049124103260341

Bollen, K.A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265-284.
doi: 10.1037/a0024448

Bollen, K.A., & Davis, W.R. (2009). Causal indicator models: identification, estimation, and testing. *Structural Equation Modeling*, *16*(3), 498–522.
doi: 10.1080/10705510903008253

Bollen, K.A., Kirby, J.B., Curran, P.J., Paxton, P.M., & Chen, F. (2007). Latent Variable Models Under Misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators. *Sociological Methods & Research*, *36*(1), 48-86. doi: 10.1177/0049124107301947

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *100*(2), 305-314.

Bollen, K.A., & Maydeu-Olivares, A. (2007). A Polychoric Instrumental Variable (PIV) Estimator for Structural Equation Models with Categorical Variables. *Psychometrika*, *72*(3), 309-326. doi: 10.1007/s11336-007-9006-3

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling*, *7*(3), 461-483.
doi: 10.1207/S15328007SEM0703_6

Boomsma, A., & Hoogland, J.J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit & D. S□orbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139-168). Lincolnwood, IL: Scientific Software International.

Borsboom, D., Mellenbergh, G.J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219.

Boslaugh, S. (2007). *Secondary Data Sources for Public Health: A Practical Guide*. New York, NY: Cambridge University Press.

Brady, H.E., Grand, S.A., & Powell, M.A., Schink, W. (2001). Access and Confidentiality Issues with Administrative Data. In M. ver Ploeg, R.A. Moffitt & C.F.Citro (Eds.), *Studies of Welfare Populations: Data Collection and Research Issues* (Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs, pp. 220-274). Washington, DC: National Academy Press.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford.

Buckler, F., & Hennig-Thurau, T. (2008). Identifying Hidden Structures in Marketing's Structural Models Through Universal Structure Modeling: An Explorative Bayesian Neural Network Complement to LISREL and PLS. *Marketing -- Journal of Research and Management, 4*(2), 47-66.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.

Cameron, A.C., Gelbach, J.B., & Miller, D.L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238-249.
doi: 10.1198/jbes.2010.07136

Chakrabarti, S., Cox, E., Frank, E., Güting, R.H., Han, J., Jiang, X., Kamber, M., Lightstone, S.S., Nadeau, T.P., Neapolitan, R.E., Pyle, D., Refaat, M., Schneider, M., Teorey, T.J., & Witten, I.H. (2009). *Data Mining: Know It All*. Burlington, MA: Morgan Kauffman.

Chen, F., Bollen, K.A., Paxton, P., Curran, P.J., & Kirby, J.B. (2001). Improper Solutions in Structural Equation Models: Causes, Consequences, and Strategies. *Sociological Methods & Research*, *29*(4), 468-508.
doi: 10.1177/0049124101029004003

Chin, W. W. (1998). The partial least squares approach to structural equation modelling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chin, W. W. (2003). *PLS-Graph user's guide, Version 3.0*. Houston, TX: Soft Modelling Inc.

Chin, W. W. (2010). Bootstrap Cross-Validation Indices for PLS Path Model Assessment. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications (pp. 83-97). Berlin Heidelberg: Springer-Verlag.

Chin, W.W., & Dibbern, J. (2010). An Introduction to a Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services Between Germany and the USA. *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 171-193). Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-540-32827-8_8

Chin, W. W., & Newsted, P. R. (1999). Structural equation modelling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Thousand Oaks, CA: Sage.

Chumney, F. L. (2013). *Structural equation models with small samples: A comparative study of four approaches*. Unpublished doctoral dissertation. Lincoln, NE: University of Nebraska. URL http://digitalcommons.unl.edu/cehsdiss/189/

Cole, D.A., Ciesla, J.A., & Steiger, J.H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent variable covariance structure analysis. *Psychological Methods*, *12*(4), 381-398.
doi: 10.1037/1082-989X.12.4.381

Cote, J.A., & Greenberg, R. (1990). Specifying measurement error In structural equation models: Are congeneric measurement models appropriate? In M.E. Goldberg, G. Gorn, & R.W. Pollay (Eds.), *Advances in Consumer Research* (vol. 17, pp. 426-433). Provo, UT: Association for Consumer Research.

Crisci, A., & D'Ambra, A. (2012). Permutation test for group comparison in PLS-PATH Modeling. *Electronic Journal of Applied Statistical Analysis*, *5*(3). doi: 10.1285/i20705948v5n3p339

Davis, W.R. (1993). The FC1 rule of identification for confirmatory factor analysis: a general sufficient condition. *Sociological Methods & Research*, *21*(4), 403-437.
doi: 10.1177/0049124193021004001

de Leeuw, E., Hox, J., & Dillman, D. (Eds). (2008). *International Handbook of Survey Methodology*. New York, NY: Erlbaum.

Diamantopoulos, A., & Temme, D. (2013). MIMIC models, formative indicators and the joys of research. *AMS Review*, *3*(3), 160-170.
doi: 10.1007/s13162-013-0050-0

Dijkstra, T.K. (2010). Latent variables and indices. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 23-46). Heidelberg: Springer-Verlag. doi: 10.1007/978-3-540-32827-8_2

Dijkstra, T.K. (2014). PLS' Janus face. *Long Range Planning*, forthcoming.

Dijkstra, T.K., & Henseler, J. (2012). *Consistent and asymptotically normal PLS-estimators for linear structural equations*. Working paper.
URL http://www.rug.nl/staff/t.k.dijkstra/research

Dijkstra, T.K., & Henseler, J. (2013). *Consistent Partial Least Squares*. 2nd VOC Conference, Antwerp, Belgium, May 31.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin*, *101*(1), 126-135.

Dolan, C. (1996). Principal component analysis using LISREL 8. *Structural Equation Modeling*, *3*(4), 307-322.
doi: 10.1080/10705519609540049

Dolan, C., Bechger, T., & Molenaar, P. (1999). Using structural equation modeling to fit models incorporating principal components. *Structural Equation Modeling*, *6*(3), 233-261.
doi: 10.1080/10705519909540132

Edwards, J.R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, *14*(2), 370-388.
doi: 10.1177/1094428110378369

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of the relationship between constructs and measures. *Psychological Methods*, *5*(2), 155-174.

Elwert, F. (2013). Graphical causal models. In S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research* (pp. 245-273). Dordrecht. The Netherlands: Springer Science+Business Media. doi: 10.1007/978-94-007-6094-3_13

Enders, C.K. (2006). Analyzing structural equation models with missing data. In G.R. Hancock & R.O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 313-342). Greenwich, CT: Information Age Publishing.

Enders, C.K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, *16*(1), 1-16.
doi: 10.1037/a0022640.

Enders, C.K., & Gottschall, A.C. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling*, *18*(1), 35-54.

doi: 10.1080/10705511.2011.532695

Esposito Vinzi, V., Chin, W.W., Henseler, J., & Wang, H. (Eds.). (2010). Handbook of Partial Least Squares: Concepts, Methods and Applications. Berlin Heidelberg: Springer Verlag. doi: 10.1007/978-3-540-32827-8

Esposito Vinzi, V., Trinchera, L., & Amato, S. (2010). PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications* (pp. 47-82). Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-540-32827-8_3

Fan, J., Guo, S., & Hao, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *74*(1), 37–65. doi: 10.1111/j.1467-9868.2011.01005.x

Fan, W. (2010). Impact of Measurement Model Modification on Structural Parameter Integrity When Measurement Model is Misspecified. *Journal of Modern Applied Statistical Methods*, *9*(1), Article 12. URL http://digitalcommons.wayne.edu/jmasm/vol9/iss1/12

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*(1), 56-83. doi: 10.1080/10705519909540119

Fattore, M., Pelagatti, M.M., & Vittadini, G. (2012). Inconsistencies of the PLS-PM approach to structural equation models with formative-reflective schemes. *Electronic Journal of Applied Statistical Analysis*, *5*(3), 333–338. doi: 10.1285/i20705948v5n3p333

Fayers P.M., & Hand, D.J. (1997). Factor analysis, causal indicators, and quality of life. *Quality of Life Research*, *6*(2), 139-150. doi: 10.1023/A:1026490117121

Fayers P.M., & Hand, D.J. (2002). Causal variables, indicator variables, and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, *165*(2), 233-261. doi: 10.1111/1467-985X.02020

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50.

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65*(1), 47-55.
doi: 10.1007/s00265-010-1038-5

Foschi, P., Belsley, D.A., & Kontoghiorghes, E.J. (2003). A comparative study of algorithms for solving seemingly unrelated regressions models. C*omputational Statistics & Data Analysis*, *44*(1-2), 3-35.
doi:10.1016/S0167-9473(03)00028-8

Freedman, D.A. (2006). On the So-Called “Huber Sandwich Estimator” and “Robust Standard Errors”. *The American Statistician*, *60*(4), 299-302.
doi: 10.1198/000313006X152207

Gaber, M.M. (Ed.). (2010). *Scientific Data Mining and Knowledge Discovery: Principles and Foundations*. Berlin Heidelberg: Springer-Verlag.
doi: 10.1007/978-3-642-02788-8

Gerbing, D.W., & Anderson, J.C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11*, 572-580.

Goldstein, H. (2011). *Multilevel Statistical Models* (4th ed.).Chichester, UK: Wiley.

Goodhue, D. L., Lewis, W., & Thompson, R.L. (2012a). Comparing PLS to regression and LISREL: A response to Marcoulides, Chin, and Saunders. *MIS Quarterly, 36*(3), 703-716.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012b). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981-1001.

Goodhue, D. L., Thompson, R.L., & Lewis, W. (2013). *Why You Shouldn't Use PLS: Four Reasons to Be Uneasy about Using PLS in Analyzing Path Models*. 46th Hawaii International Conference on System Sciences (HICSS), pp. 4739-4748. Wailea, Maui, HI:
URL http://doi.ieeecomputersociety.org/10.1109/HICSS.2013.612

Grace, J.B., & Bollen, K.A. (2008). Representing general theoretical concepts in structural equation models: the role of composite variables. *Environmental and Ecological Statistics*, *15*(2), 191-213.
doi: 10.1007/s10651-007-0047-7

Grace, J.B., Schoolmaster, D.R., Jr., Guntenspergen, G.R., Little, A.M., Mitchell, B.R., Miller, K.M., & Schweiger, E.W. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, *3*, Article 73.
URL http://dx.doi.org/10.1890/ES12-00048.1

Green, S. B., Thompson, M. S., & Babyak, M. A. (1998). A Monte Carlo investigation of methods for controlling Type I errors with specification searches in structural equation modeling. *Multivariate Behavioral Research*, *33*, 365-383.
doi: 10.1207/s15327906mbr3303_3

Green, S. B., Thompson, M. S., & Poirer, J. (1999). Exploratory analyses to improve model fit: Errors due to misspecification and a strategy to reduce their occurrence. *Structural Equation Modeling*, *6*, 113-126.
doi: 10.1080/10705519909540122

Green, S. B., & Thompson, M. S., & Poirer, J. (2001). An adjusted Bonferroni method for elimination of parameters in specification addition searches. *Structural Equation Modeling*, *8*, 18-39.
doi: 10.1207/S15328007SEM0801_2

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, *29*(4), 722-729.
doi: 10.1093/ije/29.4.722

Hancock, G. R. (1999). A sequential Scheffé-type respecification procedure for controlling Type I error in exploratory structural equation model modification. *Structural Equation Modeling*, *6*, 158-168.
doi: 10.1080/10705519909540126

Haenlein, M., & Kaplan, A.M. (2004). A beginner's guide to partial least squares (PLS) analysis. *Understanding Statistics*, *3*(4), 283-297.
doi: 10.1207/s15328031us0304_4

Hair, J.F., Jr., Ringle, C.M., & Sarstedt, M. (2011). The Use of Partial Least Squares (PLS) to Address Marketing Management Topics [special issue]. *The Journal of Marketing Theory and Practice*, *19*(2), 135-252.

Hair, J.F., Jr., Ringle, C.M., & Sarstedt, M. (2012). Analytical approaches to strategic management: Partial Least Squares modeling in strategy research [special issue]. Long Range Planning, 45(5–6), 309-484.

Hair, J.F., Jr., Ringle, C.M., & Sarstedt, M. (2013a). PLS Applications in Strategic management: Partial Least Squares Modeling in Strategy Research [special issue]. Long Range Planning, 46(1–2), 1–194.

Hair, J.F., Jr., Ringle, C.M., & Sarstedt, M. (2013b). Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. Long Range Planning, 46(1–2), 1–12.
doi: 10.1016/j.lrp.2013.01.001

Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012). Applications of partial least squares path modeling in management journals: a review of past practices and recommendations for future applications. Long Range Planning, 45(5-6), 320-340. doi:10.1016/j.lrp.2012.09.008

Hair, J.F., Sarstedt, M., Ringle, & Mena, J.A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. Journal of the Academy of Marketing Science, 40, 414–433.
doi: 10.1007/s11747-011-0261-6

Hahs-Vaughn, D. L., McWayne, C., Bulotsky-Shearer, R.J., Wen, X., & Faria, A. (2011). Methodological considerations in using complex survey data: Applied examples with the Head Start Family and Child Experiences Survey (FACES). *Evaluation Review*, *35*(3), 304-313. doi: 10.1177/0193841X11412071

Hanafi, M. (2007). PLS Path modelling: computation of latent variables with the estimation mode B. *Computational Statistics*, *22*(2), 275-292.
doi: 10.1007/s00180-007-0042-3

Hardin, J.W., & Carroll, R.J. (2003). Variance estimation for the instrumental variables approach to measurement error in generalized linear models. *The Stata Journal*, *3*(4), 342-350.
URL http://www.stata-journal.com/article.html?article=st0048

Hardin, A.M., & Chang, J.C.-J. (2013). Does existing measurement theory support the use of composite and causal indicators in information systems research? *ACM SIGMIS Database, 44*(4), 56-65.
doi: 10.1145/2544415.2544419

Harrington, D. (2009). Confirmatory factor analysis. New York: Oxford University Press.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139-164.

Hausman, J.A. (1978). Specification Tests in Econometrics. Econometrica, 46(6), 1251-1271.

Hausman, J.A. (1983). Specification and estimation of simultaneous equation models. In Z. Griliches & Intriligator, M.D. (Ed.), Handbook of Econometrics (vol I, pp. 391-448). Amsterdam: North-Holland.

Hayduk, L.A., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! one, two, three – Testing the theory in structural equation models! Personality and Individual Differences, 42(5), 841-850.
doi: 10.1016/j.paid.2006.10.001

Hayduk, L.A., Cummings, G., Stratkotter, R., Nimmo, M., Grygoryev, K., Dosman, D., Gillespie, M., Pazderka-Robinson, H., & Boadu, K. (2003). Pearl's D-Separation: One More Step

Into Causal Thinking. *Structural Equation Modeling*, *10*(2), 289-311. doi: 10.1207/S15328007SEM1002_8

Hayduk, L.A., Pazderka-Robinson, H., Cummings, G.G., Levers, M.-J. D., & Beres, M.A. (2005). Structural equation model testing and the quality of natural killer cell activity measurements. *BMC Medical Research Methodology*, *5*(1). doi: 10.1186/1471-2288-5-1

Hegdé, J. (2010). Overfitting. In N.J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 982-985). Thousand Oaks, CA: Sage.

Hedges, L.V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, *32*(2), 151-179. doi: 10.3102/1076998606298040

Hedges, L.V. (2009). Adjusting a significance test for clustering in designs with two levels of nesting. *Journal of Educational and Behavioral Statistics*, *34*(4), 464-490. doi: 10.3102/1076998609337251

Heeringa, S.G., West, B.T., & Berglund, P.A. (2010). *Applied Survey Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC.

Henseler, J., Dijkstra, T., Sarstedt, M., Ringle, C., Diamantopoulos, A., Straub, D., Ketchen, D., Hair, J., Hult, T., & Calantone, R. (2014). Common beliefs and reality about PLS: comments on Rönkkö & Evermann (2013). *Organizational Research Methods*, forthcoming.

Henseler, J., Ringle, C.M., & Sinkovics, R.R. (2009). The use of partial least squares path modeling in international marketing. In R.R. Sinkovics & P.N. Ghauri (Eds.), *Advances in International Marketing* (AIM, vol. 20, pp. 277-230). Bingley, UK: Emerald Group Publishing Limited. doi: 10.1108/S1474-7979(2009)0000020014

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, *28*(2), 565-580. doi:10.1007/s00180-012-0317-1.

Herzog, W., & Boomsma, A. (2009). Small-Sample Robust Estimators of Noncentrality-Based and Incremental Model Fit. *Structural Equation Modeling*, *16*(1), 1-27. doi: 10.1080/10705510802561279

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling*, *14*(3), 361-390. doi: 10.1080/10705510701301602

Howell, R.D., Breivik, E., & Wilcox, J.B. (2013). Formative measurement: a critical perspective. *ACM SIGMIS Database*, *44*(4), 44-55. doi: 10.1145/2544415.2544418

Hox, J.J. (2013). Multilevel regression and multilevel structural equation modeling. In T.D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (vol. 2, pp. 281-294). Oxford, UK: Oxford University Press.

Hoyle, R.H. (Ed.). (2012). *Handbook of Structural Equation Modeling*. New York, NY: Guilford Press.

Hox, J.J., & Roberts, J.K. (Eds.) (2010). *Handbook of Advanced Multilevel Analysis*. New York, NY: Taylor & Francis.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strategic Management Journal, 20(2), 195–204.
doi: 10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.3.CO;2-Z

Hulland, J., Ryan, M.J., & Rayner, R.K. (2010). Modeling Customer Satisfaction: A Comparative Performance Evaluation of Covariance Structure Analysis Versus Partial Least Squares. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications* (pp. 307-325). Heidelberg: Springer-Verlag.
doi: 10.1007/978-3-540-32827-8_15

Hwang, H. (2008). VisualGSCA 1.0 - A graphical user interface software program for generalized structured component analysis. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 111-120). Tokyo: University Academic Press.

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, *74(3)*, 517-530.
doi: 10.1007/s11336-009-9119-y

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, *69*(1), 81-99. doi: 10.1007/BF02295841

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6-23.
doi: 10.1037/a0014694.

Jackson, D. L., Voth, J., Frey, M. P. (2013). A note on sample size and solution propriety for confirmatory factor analytic models. *Structural Equation Modeling*, *20*(1), 86-97.
doi: 10.1080/10705511.2013.742388

Johnson, T.L., Ayinde, K., & Oyejola, B.A. (2010). Effect of equation identification status and correlation between error terms on estimators of a system of simultaneous equations model. *Electronic Journal of Applied Statistical Analysis*, *3*(2), 115–125.
doi: 10.1285/i20705948v3n2p115

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*(2), 183–202.
doi: 10.1007/BF02289343

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Jöreskog, K.G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: historical and comparative aspects. In H. Wold & K. Jöreskog (Eds.), *Systems Under Indirect Observation: Causality, Structure, Prediction* (vol. I, pp. 263-270), Amsterdam: North-Holland.

Jung, S. (2013). Structural equation modeling with small sample sizes using two-stage ridge least-squares estimation. *Behavior Research Methods*, *45*(1), 75-81. doi: 10.3758/s13428-012-0206-0

Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions* (2nd ed.). Newbury Park, CA: Sage.

Kaplan, D., Kim, J., & Kim, S. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. Millsap, & A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology* (pp. 592-613). London, UK: Sage. doi: http://dx.doi.org/10.4135/9780857020994.n24

Kargupta, H., Han, J., Yu, P.S., Motwani, R., & Kuma, V. (2009). *Next Generation of Data Mining*. Boca Raton, FL: Chapman & Hall/CRC.

Kelley, K. & Cheng, Y. (2012). Estimation and confidence interval formation for reliability coefficients of homogeneous measurement instruments. *Methodology*, *8*(2), 39-50.doi: 10.1027/1614-2241/a000036

King, G., & Roberts, M.E. (2013). *How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About it*. Working Paper.
URL http://gking.harvard.edu/files/gking/files/robust.pdf

Kirby, J.B., & Bollen, K.A. (2009). Using Instrumental Variable (IV) Tests to Evaluate Model Specification in Latent Variable Structural Equation Models. *Sociological Methodology*, *39*(1), 327–355.
doi: 10.1111/j.1467-9531.2009.01217.x

Kline, R. B. (2013a). Reverse arrow dynamics: Feedback loops and formative measurement. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 39-76). Greenwich, CT: Information Age Publishing.

Kline, R. B. (2013b). *Principles and practice of structural equation modeling* (5th ed.). New York: Guilford Press.

Kock, N. (2013). *WarpPLS 4.0 User Manual*. Laredo, TX: ScriptWarp Systems.

Kolenikov, S. (2010). Resampling variance estimation for complex survey data. *The Stata Journal*, *10*(2), 165-199.
URL http://www.stata-journal.com/article.html?article=st0187

Kolenikov, S. (2011). Biases of parameter estimates in misspecified structural equation models. *Sociological Methodology*, *41*(1), 119–157.
doi: 10.1111/j.1467-9531.2011.01236.x

Kolenikov, S., & Bollen, K.A. (2012). Testing Negative Error Variances Is a Heywood Case a Symptom of Misspecification? *Sociological Methods & Research*, *41*(1), 124-167.
doi: 10.1177/0049124112442138

Kontoghiorghes, E.J., & Dinenis, E. (1997). Computing 3SLS solutions of simultaneous equation models with a possible singular variance-covariance matrix. *Computational Economics*, *10*(3), 231-250.
doi: 10.1023/A:1008617207791

Kovacevic, M.S., Rong, H., & You, Y. (2006). *Bootstrapping for variance estimation in multi-level models fitted to survey data*. Proceedings of the Survey Research Methods Section, ASA (2006). Montréal, Quebec, Canada.

Kubáček, L. (2013). Seemingly unrelated regression models. *Applications of Mathematics*, *58*(1), 111-123.
doi: 10.1007/s10492-013-0005-7

Kyono, T.M. (2010). *Commentator: A Front-End User-Interface Module for Graphical and Structural Equation Modeling*. Technical Report R-364. Los Angeles, CA: Department of Computer Science, UCLA.
URL http://ftp.cs.ucla.edu/pub/stat_ser/r364.pdf

MacKenzie, S.B., Podsakoff, P.M., & Jarvis, C.B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710–730.

Landsheer, J.A. (2010). The specification of causal models with TETRAD IV: a review. *Structural Equation Modeling*, *17*(4), 703-711.
doi:10.1080/10705511.2010.510074

Lee, Y.-t., & Antonakis, J. (2012). When Preference Is Not Satisfied but the Individual Is: How Power Distance Moderates Person–Job Fit. *Journal of Management*, Online first.
doi: 10.1177/0149206311436080

Lee, S.-Y., , & Song, X.-Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. Chichester, UK: Wiley.

Lehtonen, R., & Pahkinen, E. (2004). *Practical Methods for Design and Analysis of Complex Surveys* (2nd ed.). Chichester, UK: Wiley.

Lei, P-W., & Wu, Q. (2012). Estimation in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 164–179). New York, NY: Guilford Press.

Lin, T.Y., Xie, Y., Wasilewska, A., & Liau, C.-J. (Eds.). (2008). *Data Mining: Foundations and Practice*. Berlin Heidelberg: Springer-Verlag.

Little, T.D., Rhemtulla, M., Gibson, K., & Schoemann, A.M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285-300. doi: 10.1037/a0033266

Löfstedt, T., Eriksson, L., Wormbs, G., & Trygg, J. (2012). Bi-modal OnPLS. *Journal of Chemometrics*, *26*(6), 236–245.
doi: 10.1002/cem.2448

Löfstedt, T., Hanafi, M., Mazerolles, G., & Trygg, J. (2012). OnPLS path modelling. *Chemometrics and Intelligent*, *118*, 139–149.
doi: 10.1016/j.chemolab.2012.08.009

Löfstedt, T., Hanafi, M., & Trygg, J. (2013). Multiblock and Path Modeling with OnPLS. In H. Abdi, W.W. Chin, V. Esposito Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New Perspectives in Partial Least Squares and Related Method* (vol. 56, pp. 209-220). New York, NY: Springer Science+Business Media. doi: 10.1007/978-1-4614-8283-3_14

Löfstedt, T., Hoffman, D., & Trygg, J. (2013). Global, local and unique decompositions in OnPLS for multiblock data analysis. *Analytica Chimica Acta*, *791*, 13-24. doi: 10.1016/j.aca.2013.06.026

Löfstedt, T., & Trygg, J. (2011). OnPLS—a novel multiblock method for the modelling of predictive and orthogonal variation. *Journal of Chemometrics*, *25*(8), 441–455. doi: 10.1002/cem.1388

Lohmöller, J.-B. (1984). *LVPLS Program Manual: Latent Variables Path Analysis with Partial Least-Squares Estimation*. Köln: Zentralarchiv für empirische Sozialforschung.

Lohmöller, J.-B. (1988). The PLS program system: Latent variables path analysis with partial least squares estimation. *Multivariate Behavioral Research*, *23*(1), 125-127. doi:10.1207/s15327906mbr2301_7

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

Lovaglio, P.G., & Vittadini, G. (2013). Component Analysis for Structural Equation Models with Concomitant Indicators In P. Giudici, S. Ingrassia, & M. Vichi (Eds.), *Statistical Models for Data Analysis: Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 209-215). Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-00032-9_24

Lumley, T. (2011). *Complex Surveys: A Guide to Analysis Using R*. Hoboken, NJ: Erlbaum.

Maas, C.J.M., & Hox, J.J. (2004a). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*, 427–440.

Maas, C.J.M., & Hox, J.J. (2004b). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*(2), 127–137.
doi: 10.1046/j.0039-0402.2003.00252.x

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84-99.

Marchetti, G.M., & Drton, M. (2013). *Graphical Gaussian Models: Package 'ggm', Version 1.995-3*.
URL http://cran.r-project.org/web/packages/ggm/

Marcoulides, G.A., & Chin, W.W. (2013). You write, but others read: common methodological misunderstandings in PLS and related methods. In H. Abdi, W.W. Chin, V. Esposito Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New Perspectives in Partial Least Squares and Related Methods* (pp. 31-64). New York, NY: Springer Science+Business Media. doi: 10.1007/978-1-4614-8283-3_2

Marcoulides, G.A., Chin, W.W., & Saunders, C.S. (2012). When imprecise statistical statements become problematic: a response to Goodhue, Lewis, and Thompson. *MIS Quarterly*, *36*(3), 717-728.

Marcoulides, G.A., & Ing, M. (2012). Automated structural equation modeling strategies. In R. Hoyle (Ed.). *Handbook of structural equation modeling* (pp. 690-704). New York, NY: Guilford Press.

Marcoulides, G.A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet? *MIS Quarterly, 30*(2), iii-ix.

Marsh, H. W., Hau, K-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McCardle (Eds.), *Contemporary psychometrics: A Festschrift to Roderick P. McDonald* (pp. 275-340). Mahwah NJ: Erlbaum.

Marsh, H.W., Lüdtke, O., Nagengast, B., Morin, A.J.S., Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—camouflaging misspecification with item parcels in CFA. models. *Psychological Methods*, *18*(3), 257-284. doi: 10.1037/a0032773

Martin, M. & Roberts, S. (2010). Data Snooping. In N.J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 323-333). Thousand Oaks, CA: Sage.
doi: http://dx.doi.org/10.4135/9781412961288

Mateos-Aparicio, G. (2011). Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. *Communications in Statistics - Theory and Methods*, *40*(13), 2305-2317. doi:10.1080/03610921003778225

Maydeu-Olivares, A., Coffman, D. L., García-Forero, C. & Gallardo-Pujol, D. (2010). Hypothesis testing for coefficient alpha: An SEM approach. *Behavior Research Methods*, *42*, 618-625.
doi: 10.3758/BRM.42.2.618

McCoach, D.B., & Adelson, J.L. (2010). Dealing with dependence (Part I): understanding the effects of clustered data. *Gifted Child Quarterly*, 54(2), 152–155.
doi: 10.1177/0016986210363076

McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, *31*(2), 239-270. doi:10.1207/s15327906mbr3102_5

McIntosh, C. N. (2007). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett. *Personality and Individual Differences*, *42*(5), 859–867. doi:10.1016/j.paid.2006.09.020

McIntosh, C.N. (2012). The presence of an error term does not preclude causal inference in regression: a comment on Krause (2012). *Quality & Quantity*, Online first.
doi: 10.1007/s11135-012-9763-7

McIntosh, C.N. (2013). Strengthening the assessment of factorial invariance across population subgroups: a commentary on Varni et al. (2013). *Quality of Life Research*, *22*(9), 2595-2601.
doi: 10.1007/s11136-013-0465-y

Monecke, A. (2013). *Structural equation modeling using partial least squares: Package ‘semPLS’, Version 1.0-10*. URL http://cran.r-project.org/web/packages/semPLS/index.html

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313-335.
doi: 10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2012). *Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation--Recommendations to the National Center for Education Statistics*. Jessup, MD: ED Pubs. URL http://eric.ed.gov/?id=ED517134

Nestler, S. (2013a). How the 2SLS/IV estimator can handle equality constraints in structural equation models: A system-of-equations approach. *British Journal of Mathematical and Statistical Psychology*, Early View.
doi: 10.1111/bmsp.12023

Nestler, S. (2013b). A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology, 66*(1), 127–143.
doi: 10.1111/j.2044-8317.2012.02044.x

Nunnally, J. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd Ed.). New York: McGraw-Hill.

Oberski, D.L. (2013). lavaan.survey: An R Package for Complex Survey Analysis of Structural Equation Models. *Journal of Statistical Software*, forthcoming.
URL http://daob.nl/wp-content/uploads/2013/05/lavaan-survey.pdf

Osborne, J. (2011). Best practices in using large, complex samples: the importance of using appropriate weights and designe Effect compensation. *Practical Assessment, Research & Evaluation*, *16*(12). URL
http://pareonline.net/getvn.asp?v=16&n=12.

Padilla, M.A., & Divers, J. (2013). Coefficient Omega Bootstrap Confidence Intervals: Nonnormal Distributions. *Educational and Psychological Measurement*, *73*(6), 956-972.
doi: 10.1177/0013164413492765

Pal, S. (2009). Rao and Wu's re-scaling bootstrap modified to achieve extended coverages. *Journal of Statistical Planning and Inference*, *139*(10), 3552-3558.
doi: 10.1016/j.jspi.2009.04.006

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press.

Peng, D.X., & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, *30*(6), 467–480.

doi: 10.1016/j.jom.2012.06.002

Pesarin, F., & Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. Chichester, UK: Wiley.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, *61*(2), 317-337.

Pfeffermann, D. (1996). The use of sampling weights for survey data analysis. *Statistical Methods in Medical Research*, 5(3), 239–261. doi: 10.1177/096228029600500303

Pierre, F., & Saidi, A. (2008). *Implementing Resampling Methods for Design-Based Variance Estimation in Multilevel Models: Using HLM6 and SAS Together*. Proceedings of the Survey Research Methods Section (2008), Joint Statistical Meeting, Denver, Colorado.
URL http://www.amstat.org/sections/srms/proceedings/y2008/Files/300724.pdf

Preston, J. (2009). Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, *35*(2), 227-234. Statistics Canada Catalogue No. 12-001-X.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL http://www.R-project.org

Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2007). Multilevel structural equation modeling. In S-Y. Lee (Ed.), *Handbook of Latent Variable and Related Models* (pp. 209-227). Amsterdam: Elsevier.

Raghav, M., & Barreto, H. (2011). *Understanding and teaching unequal probability of selection*. MPRA Paper 32334. Munich, Germany: University Library of Munich.
URL http://ideas.repec.org/p/pra/mprapa/32334.html

Ranalli, M.G., & Mecatti, F. (2012). *Comparing Recent Approaches For Bootstrapping Sample Survey Data: A First Step Towards A Unified Approach*. Proceedings of the Section on Survey Research Methods – Joint Statistical Meetings 2012, San Diego, CA.
URL http://www.amstat.org/sections/srms/proceedings/y2012/files/304787_73724.pdf

Ratner, B. (2012). *Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data* (2nd ed.). Boca Raton, FL: CRC Press, Taylor & Francis.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173-184.
doi: 10.1177/01466216970212006

Raykov, T. (2012). Evaluation of latent construct correlations in the presence of missing data: A note on a latent variable modelling approach. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 19–31.
doi: 10.1348/000711010X498162

Raykov, T., & Marcoulides, G.A. (2010). *Introduction to Psychometric Theory*. New York, NY: Taylor & Francis.

Rai, A., Goodhue, D., Henseler, J., & Thompson, R. (2013). *To PLS or not to PLS: that is the question*. 19th Americas Conference on Information Systems. Chicago, Illinois.
URL http://aisel.aisnet.org/amcis2013/Panels/PanelSubmissions/2/

Reddy, S.K. (1992). Effects of Ignoring Correlated Measurement Error in Structural Equation Models. *Educational and Psychological Measurement*, *52*(3), 549-570.
doi: 10.1177/0013164492052003005

Rigdon, E.E. (2012). Rethinking partial least squares path modeling: in praise of simple methods. *Long Range Planning*, *45*(5–6), 341–358.
doi: 10.1016/j.lrp.2012.09.010

Ringle, C.M., Sarstedt, M., & Straub, D.W. (2012). Editor's comments: a critical look at the use of PLS-SEM in MIS quarterly. *MIS Quarterly*, *36*(1), iii-xiv.

Ringle, C.M., Wende, S., & Will, A. (2005). *SmartPLS 2.0 (M3).* Hamburg, Germany. URL http://www.smartpls.de.

Roberts, J. K. & Fan, X. (2004). Bootstrapping within the multilevel/hierarchical linear modeling framework: A primer for use with SAS and S-PLUS. *Multiple Linear Regression Viewpoints*, *30*(1), 23-34. URL http://mlrv.ua.edu/2004/vol30_N1_A5.pdf

Rocha, C.M., & Chelladurai, P. (2012). Item parcels in structural equation modeling: an applied study in sport management. *International Journal of Psychology and Behavioral Sciences*, *2*(1), 46-53. doi: 10.5923/j.ijpbs.20120201.07

Rodgers, J.L., & Beasley, W.H. (2013). Fisher, Gosset, and alternative hypothesis significance testing (AHST): using the bootstrap to test scientific hypotheses about the multiple correlation. In M.C Edwards & R.C. MacCallum, (Eds.), *Current Topics in the Theory and Application of Latent Variable Models* (pp. 217-239). New York, NY: Routledge.

Ronchetti, E., Field, C., & Blanchard, W. (1997). Robust Linear Model Selection by Cross-Validation. *Journal of the American Statistical Association*, *92*(439), 1017-1023
doi: 10.1080/01621459.1997.10474057

Rönkkö, M. (2014). The effects of chance correlations on partial least squares path modeling. *Organizational Research Methods*, forthcoming.

Rönkkö, M. (2013). *Matrix-based partial least squares estimation: Package 'matrixpls', Version 0.1.0*. URL https://github.com/mronkko/matrixpls

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425-448. doi: 10.1177/1094428112474693

Rönkkö, M., & Ylitalo, J. (2010). *Construct Validity in Partial Least Squares Path Modeling*. 2010 International Conference on Information Systems (ICIS) (Paper 155). St. Louis: AIS. URL http://aisel.aisnet.org/icis2010_submissions/155

Sanchez, G., & Trinchera, L. (2013). *Partial least squares data analysis methods: package 'plspm', Version 0.1-11*. URL http://cran.r-project.org/web/packages/plspm/index.html

Saris, W. E., & Aalberts, C. (2003). Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling*, *10*(2), 193-213. doi: :10.1207/S15328007SEM1002_2

Saris, W.E., Satorra, A., & van der Veld, W. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*(4), 561-582. doi: 10.1080/10705510903203433

Savalei, V. (in press). Understanding Robust Corrections in Structural Equation Modeling. *Structural Equation Modeling*.

Savalei, V. (2010a). Small sample statistics for incomplete nonnormal data: extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling*, *17*(2), 245–268. doi: 10.1080/10705511003659375

Savalei, V. (2010b). Expected vs. observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, *15*, 352-367. doi: doi: 10.1037/a0020143

Savalei, V., & Falk, C. (in press). Robust two-stage approach outperforms robust FIML with incomplete nonnormal data. *Structural Equation Modeling*.

Savalei, V., & Yuan, K.-H. (2009). On the model-based bootstrap with missing data: obtaining a p-value for a test of exact fit. *Multivariate Behavioral Research*, *44*(6), 741-763. doi: 10.1080/00273170903333590

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 of Psychological Research-Online*, *8*(2), 23-74. URL http://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Hillsdale, NJ: Erlbaum.

Seco, G.V., García, M.A., García, M.P.F., & Rojas, P.E.L. (2013). Multilevel bootstrap analysis with assumptions violated. *Psicothema*, *25*(4), 520-528. doi: 10.7334/psicothema2013.58

Semadeni, M., Withers, M.C., & Certo, S.T. (2013). The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal*, Early View. doi: DOI: 10.1002/smj.2136

Semykina, A. (2012). Specification tests and tests for overidentifying restrictions in panel data models with selection. *Economics Letters*, *15*(1). 53-55. doi:10.1016/j.econlet.2011.11.019

Sharma, P.N., & Kim, K. H. (2013). A Comparison of PLS and ML Bootstrapping Techniques in SEM: A Monte Carlo Study. In H. Abdi, W.W. Chin, V. Esposito Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New Perspectives in Partial Least Squares and Related Methods* (pp. 201-208). New York, NY: Springer Science+Business Media. doi: 10.1007/978-1-4614-8283-3_13

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A, Kawahara, Y., Washio, T., Hoyer, P.O., & Bollen, K.A. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, *12*, 1225-1248. URL http://jmlr.csail.mit.edu/papers/volume12/shimizu11a/shimizu11a.pdf

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, *7*, 206-218. doi: 10.1207/S15328007SEM0702_4

Shipley, B. (2002). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference* (2nd ed). Cambridge, UK: Cambridge University Press.

Shipley, B. (2003). Testing recursive path models with correlated errors using d-separation. *Structural Equation Modeling*, *10*, 214-221. doi:10.1207/S15328007SEM1002_3

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Snijders, T.A.B., & Bosker, R. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Spirtes, P., Glymour, C., Scheines, R., & Tillman, R. (2010). Automated Search for Causal Relations: Theory and Practice. In R. Dechter, H. Geffner, J.Y. Halpern (Eds.), *Heuristics, Probability, and Causality: A Tribute to Judea Pearl* (pp. 467-506). London, UK: College Publications.

Stapleton, L. M. (2008). Variance estimation using replication methods in structural equation modeling with complex sample data. *Structural Equation Modeling*, *15*(2), 183-210.

Steiger, J.H., Shapiro, A., & Browne, M.W. (1985). On the multivariate distribution of sequential chi-square statistics. *Psychometrika*, *50*(3), 253-264.
doi: 10.1007/BF02294104

Sterba, S.K. (2011). Implications of Parcel-Allocation Variability for Comparing Fit of Item-Solutions and Parcel-Solutions. *Structural Equation Modeling*, *18*(4), 554-577.
doi: 10.1080/10705511.2011.607073

Sterba, S.K., & MacCallum, R.C. (2010). Variability in Parameter Estimates and Model Fit Across Repeated Allocations of Items to Parcels. *Multivariate Behavioral Research*, *45*(2), 322-358.
doi: 10.1080/00273171003680302

Subramanian, J., & Simon, R. (2013). Overfitting in prediction models - Is it a problem only in high dimensions? *Contemporary Clinical Trials*, *36*(2), 636-641.
doi: 10.1016/j.cct.2013.06.011

Swain, A. J. (1975). *Analysis of parametric structures for variance matrices*. Doctoral dissertation. Adelaide: University of Adelaide.
URL http://ebooks.adelaide.edu.au/dspace/handle/2440/21286

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management and Business Excellence*, *19*, 871-886.
doi: 10.1080/14783360802159543

Tenenhaus, A. (2013). *Regularized Generalized Canonical Correlation Analysis: Package 'RGCCA', Version 2.0.*
URL http://cran.r-project.org/web/packages/RGCCA/index.html

Tenenhaus M., Esposito, V.V., Chatelin Y.M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, *48*(1), 159-205.

Tenenhaus, M., Mauger, E., & Guinot, C. (2010). Use of ULS-SEM and PLS-SEM to Measure a Group Effect in a Regression Model Relating Two Blocks of Binary Variables. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang, (Eds.), Handbook of Partial Least Squares: Concepts, Methods, and Applications (pp. 125-140). Heidelberg: Springer-Verlag.
doi: 10.1007/978-3-540-32827-8_6

Tenenhaus, A., & Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika*, *76*(2), 257-284.
doi: 10.1007/s11336-011-9206-8

Textor, J. (2013). *Drawing and Analyzing Causal DAGs with DAGitty: User Manual for Version 2.0*.
URL http://www.dagitty.net/manual-2.x.pdf

Thompson, S.B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, *99*(1), 1–10.
doi: 10.1016/j.jfineco.2010.08.016

Treiblmaier, H., Bentler, P.M., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling*, *18*(1), 1-17.
doi: 10.1080/10705511.2011.532693

Trujillo. G.S. (2009). *PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling*. Barcelona: Universitat Politècnica de Catalunya.

Trzesniewski, K.H., Donnellan, M.B., & Lucas, R.E. (Eds.). (2011). *Secondary Data Analysis: An Introduction for Psychologists*. Washington, DC: American Psychological Association.

Tu, S., & Xu, L. (2011). Parameterizations make different model selections: Empirical findings from factor analysis. *Frontiers of Electrical and Electronic Engineering in China*, *6*(2), 256–274.
URL http://www.cse.cuhk.edu.hk/~lxu/papers/journal/11FEE-tsk-two.pdf

Turkyilmaz, A., Oztekin, A., Zaim, S., & Demirel, O.F. (2013). Universal structure modeling approach to customer satisfaction index. *Industrial Management & Data Systems*, *113*(7), 932-949.
doi: 10.1108/IMDS-12-2012-0444

Ullman, J.B. & Bentler, P.M. (2003). Structural equation modeling. In Weiner I.B., Schinka J.A., & Velicer W.F. (Eds.), *Handbook of Psychology: Research methods in Psychology* (Vol. 2, pp. 607-634). New York: Wiley.

Vartanian, T.P. (2011). *Secondary Data Analysis*. New York, NY: Oxford University Press.

Vidaurre, D., van Gerven, M.A., Bielza, C., Larrañaga, P., & Heskes, T. (2013). Bayesian sparse partial least squares. *Neural Computation*, *25*(12), 3318-3339.
doi: 10.1162/NECO_a_00524

Vilares, M.J., Almeida, M.H., & Coelho, P.S. (2010). Comparison of Likelihood and PLS Estimators for Structural Equation Modeling: A Simulation with Customer Satisfaction Data. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications* (pp. 289-305). Heidelberg: Springer-Verlag. doi: 10.1007/978-3-540-32827-8_14

Vilares, M.J., & Coelho, P.S. (2013). Likelihood and PLS Estimators for Structural Equation Modeling: An Assessment of Sample Size, Skewness and Model Misspecification Effects. In J.L. da Silva, F. Caeiro, I. Natário, & C.A. Braumann (Eds.), *Advances in Regression, Survival*

Analysis, Extreme Values, Markov Processes and Other Statistical Applications (pp. 11-33). Berlin Heidelberg: Springer-Verlag.
doi: 10.1007/978-3-642-34904-1_2

Vittadini, G., Minotti, S.C., Fattore, M., & Lovaglio, P.G. (2007). On the relationships among latent variables and residuals in PLS path modeling: The formative-reflective scheme. *Computational Statistics & Data Analysis*, *51*, 5828– 846.
doi: 10.1016/j.csda.2006.10.023

Wang, J., & Wang, X. (2012). *Structural Equation Modeling: Applications Using Mplus*. Chichester, UK: Wiley.

Westfall, P.H., Henning, K.S.S., & Howell, R.D. (2012). The Effect of Error Correlation on Interfactor Correlation in Psychometric Measurement. Structural Equation Modeling, 19(1), 99-117.
doi :10.1080/10705511.2012.634726

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.), *Multivariate Analysis II* (pp.391-420). New York: Academic Press.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modeling: some current developments. In P.R. Krishnaiah (Ed.), *Multivariate Analysis II* (pp. 383-407). New York: Academic Press,

Wold, H. (1975). Path models with latent variables: The NIPALS Approach. In H.M. Blalock, A. Aganbegian, F.M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative Sociology* (pp. 307-359). New York: Academic Press.

Wold, H. (1982). Soft modeling: the basic design and some extensions. In J.K. Jöreskog, & H. Wold (Eds.), *Systems under Indirect Observation* (Pt. 2, pp. 1-54). Amsterdam: North-Holland.

Wold, H. (1985). Partial least squares. In S. Kotz & N.L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (vol. 6, pp. 581-591). New York, NY: Wiley.

Wold, S., Eriksson, L., & Kettaneh, N. (2010). PLS in data mining and data integration. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang, (Eds.), *Handbook of Partial Least Squares: Concepts, Methods, and Applications*, pp. 327-357). Heidelberg: Springer-Verlag.
doi: 10.1007/978-3-540-32827-8_16

Wolf, E.J., Harrington, K.M., Clark, S.L., & Miller, M.W. (2013). Sample Size Requirements for Structural Equation Models An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913-934.
doi: 10.1177/0013164413495237

Wu, J.-J., & Kwok, O.-M. (2012). Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches. *Structural Equation Modeling*, *19*(1), 16-35.
doi: 10.1080/10705511.2012.634703

Xu, L. (2010). Bayesian Ying-Yang system, best harmony learning, and five action circling. *Frontiers of Electrical and Electronic Engineering in China*, *5*(3), 281–328.
URL http://www.cse.cuhk.edu.hk/~lxu/papers/journal/byy10.pdf

Xu, L. (2012). On essential topics of BYY harmony learning: Current status, challenging issues, and gene analysis applications. *Frontiers in Electrical and Electronic Engineering*, *7*(1), 147–196.
URL http://www.cse.cuhk.edu.hk/~lxu/papers/journal/byy12.pdf

Yetkiner Z. E., & Thompson B. (2010). Demonstration of how score reliability is integrated into SEM and how reliability affects all statistical analyses. *Multiple Linear Regression Viewpoints*, *26*(2), 1–12.
URl http://mlrv.ua.edu/2010/vol36_2/Yetkiner_proof.pdf

Yuan, K.-H., & Bentler, P.M. (2004). On chi-square difference and z-tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*, *64*, 737-757.
doi: 10.1177/0013164404264853

Yuan, K.-H., & Chan, W. (2008). Structural equation modeling with near singular covariance matrices. *Computational Statistics & Data Analysis*, *52*(10), 4842-4858.
doi: 10.1016/j.csda.2008.03.030

Yuan, K-H., Kouros, C. D., & Kelley, K. (2008). Diagnosis for covariance structure models by analyzing the path. *Structural Equation Modeling*, *15*(4), 564-602.
doi: 10.1080/10705510802338991

Yuan, K-H., Marshall, L.L., & Bentler, P.M. (2003). Assessing the Effect of Model Misspecifications on Parameter Estimates in Structural Equation Models. *Sociological Methodology*, *33*(1), 241-265.
doi: 10.1111/j.0081-1750.2003.00132.x

Yuan, K.H., Wu, R., & Bentler, P.M. (2010). Ridge structural equation modelling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology*, *64*(1), 107–133.
doi: 10.1348/000711010X497442

Yuan, K.H., & Zhang, Z. (2012). Robust Structural Equation Modeling with Missing Data and Auxiliary Variables. *Psychometrika*, *77*(4), 803-826.
doi: 10.1007/s11336-012-9282-4

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348-368. doi: 10.1080/01621459.1962.10480664

Zellner, A. (1963). Estimators for seemingly unrelated regression equations: Some exact finite sample results. *Journal of the American Statistical Association*, *58*(304), 977-992. doi: 10.1080/01621459.1963.10480681

Zheng, Z.E., & Pavlou, P.A. (2010). Research Note—Toward a Causal Interpretation from Observational Data: A New Bayesian Networks Method for Structural Models with Latent Variables. *Information Systems Research*, *21*(2), 365-391. doi: 10.1287/isre.1080.0224

Table 1. Summary of positions on key issues related to application of PLS-PM

Issue	Rönkkö and Evermann's critique	Henseler et al.'s response	McIntosh et al.'s reflection
1. Can PLS-PM be characterized as an SEM method?	***No.*** PLS-PM yields biased parameter estimates, offers no model overidentification tests, and cannot correct for endogeneity in predictors.	***Yes.*** PLS-PM is a version of SEM designed to analyze relationships among composite rather than common factors. Bias is only an artifact arising from incorrect interpretation of PLS-PM estimates as common factor model estimates. Global overidentification tests and supplemental indices are available to evaluate model fit.	***No.*** PLS-PM is not currently able to adequately estimate and test causal structures. Local tests of constrained parameters are needed to supplement global fit assessments in PLS-PM. Whereas tests of hypothesized null relationships are possible in PLS-PM, current software limitations prevent the use of equality constraints, which must instead be implemented using SEM methods for composite variables. Instrumental variable methods are required for removing endogeneity bias. Correlations between endogenous disturbances must also be taken into account.
2. Can PLS-PM reduce the impact of measurement error?	***No.*** PLS-PM cannot accommodate measurement error and is particularly sensitive to unmodeled correlated errors, which result in biased estimates.	***Yes.*** PLS-PM capitalizes on correlations among composite factor indicators in order to produce more reliable constructs. A weighted PLS-PM composite will be more reliable than an unweighted sum, provided that item reliabilities are heterogeneous, sample size is sufficiently large, and the composites are moderately correlated.	***No.*** Under the common factor model, SEM is clearly superior to PLS-PM, because it purges measurement error from the indicators, and correlated errors can be modeled. Under the composite factor model, both PLS-PM and reparameterized versions of SEM are appropriate, but measurement error will always be an inherent part of the composites, even with large numbers of highly correlated indicators. The instrumental variables approach is needed to help achieve consistent estimation in the presence of measurement error.
3. Is PLS-PM capable of validating measurement models?	***No.*** PLS-PM cannot validate measurement models, as the various indices available for evaluating model fit (e.g., composite reliability, average	***Yes.*** Rönkkö and Evermann relied on the common rather than composite factor model, made several mistakes in computing and reporting the various fit	***Yes, in the composite factor case.*** Given that measurement error can never be completely eliminated from composite variables, PLS-PM is unsuitable for validating common factor-based measurement models, and is limited to the composite case. Researchers need to maintain a distinction between two

	variance extracted, relative goodness-of-fit index) cannot reliably detect different types of misspecifications.	measures, and did not explicitly compare PLS-PM with SEM. A revised version of the analysis demonstrated that the chi-square test and the standardized root mean squared residual are able to consistently detect different type of misspecifications, and performed better when using SEM. However, high rates of non-convergence and improper solutions in SEM demonstrated the overall superiority of PLS-PM, which always converged.	types of overall model assessment criteria: (1) measures of explained variance (e.g., CR, AVE); and (2) model fit statistics (e.g., chi-square, SRMR). Under the common factor model, other, more detailed evaluation criteria include item loadings, measurement errors, and fitted residuals. Under the composite factor model, researchers should recast the detailed assessment of measurement quality in terms of the predictive ability of the composites, which should be further verified by cross-validation. Increased rates of non-convergence and improper solutions in SEM are typically observed under model misspecification, which supports rather than refutes the ability of SEM to detect incorrect models. PLS-PM's unfailing convergence may result in increased acceptance of bad models.
4. Does PLS-PM provide valid inference on path coefficients?	***No.*** PLS-PM's reliance on *t*-tests for coefficient inference is compromised by: (1) bimodal distributions for parameter estimates under the null hypothesis; and (2) discrepancies between analytical and bootstrap sampling distributions.	***Yes.*** PLS-PM path coefficients will not be bimodally distributed in all situations. Unimodality can be achieved with larger sample sizes and stronger effect sizes. A variety of bootstrapped-based approaches for building confidence intervals can offset the impact of unimodality and discrepancies between analytical and bootstrap sampling distributions.	***Inconclusive.*** Further research is required to determine the behavior of PLS-PM path coefficients under the null. Researchers can also bypass rigid distributional assumptions by using permutation (randomization) tests, which have yet to be adopted for regular use in PLS-PM.
5. Is PLS-PM more advantageous than SEM at small sample sizes?	***No.*** Any apparent advantages of PLS-PM over SEM at small sample sizes can be attributed to: (1) ignoring the effects of chance correlations among measurement errors, which inflates parameter estimates;	***Yes***. Interpretation of the small sample performance of PLS-PM relative to SEM is often confounded by the differences between composite and common factor models. However, the unfailing convergence of PLS-PM	***Inconclusive.*** Extant simulation research comparing PLS-PM and SEM has neglected several methodological innovations designed to improve SEM's performance in the small sample case (e.g., Swain corrections, ridge regressions adjustments, new ML search algorithms). Future comparative work needs to incorporate these advances in order to better determine the

	and (2) the use of a *t* distribution for NHST when the coefficient distributions are not normal, which leads to increased Type I error rates.	at small samples is a definite advantage over SEM.	relative small sample performance of PLS-PM and SEM.
6. Can PLS-PM be used for exploratory modeling?	***No.*** PLS-PM is incapable of "discovering" causal models, as it requires researchers to provide a complete model specification based on theory. It also does not offer the same model modification and exploration capabilities as SEM.	***Yes.*** PLS-PM can be seen as exploratory alternative to SEM when the common factor model does not fit well. One can also start with a fully saturated PLS-PM model and remove the non-significant pathways in a second step. SEM modification indices are not reliable and often capitalize on chance.	***Yes, but options are currently limited.*** Researchers should decide *a priori* whether the common or composite factor model is more appropriate based on careful consideration of the nature of the observed indicators and their most likely relationship with the construct (i.e., formative versus reflective). The strategy of specifying a saturated PLS-PM model and then removing non-significant paths cannot be routinely trusted to lead to the correct model. A principled exploratory approach called *universal structure modeling* is available in PLS-PM. Future work should focus on developing additional automated model search procedures appropriate for the composite factor situation. Ideas could be taken from the SEM field, which offers a number of exploratory model search algorithms.
7. Conclusion: Should PLS-PM be abandoned?	***Yes.*** PLS-PM should be abandoned in favor of SEM, which has more powerful, versatile, tried-and-true methods of parameter estimation and model testing. Even regression with either summed scales or factor scores is preferable to PLS-PM. PLS-PM is perhaps appropriate when the only research goal is prediction, but there is a lack of supporting evidence.	***No.*** PLS-PM has survived Rönkkö and Evermann's critiques. PLS-PM should remain in the statistical toolkits of applied researchers. Future research should focus on the performance of PLS-PM in estimating composite factor models and predicting outcomes.	***Yes, in all models involving common factors.*** PLS-PM should be completely abandoned for use in both common factor measurement models and path models with common factors, as it cannot compete with SEM's capabilities in this area. PLS-PM could still be used with composite factor measurement models and path models with composite variables, but several recommendations for improvement in model estimation and testing need to be incorporated (e.g., localized model fit testing, instrumental variables estimation, and techniques for dealing with correlated prediction errors). In future work, PLS-PM and SEM should be more thoroughly compared in terms of their ability to analyze composite factor models and predict outcomes.

Table 2. Reliability as a function of the average interitem correlation and the number of items that constitute a scale.

	Average Interitem Correlation								
# of Items	.10	.20	.30	.40	.50	.60	.70	.80	.90
2	.18	.33	.46	.57	.67	.75	.82	.89	.95
3	.25	.43	.56	.67	.75	.82	.88	.92	.96
4	.31	.50	.63	.73	.80	.86	.90	.94	.97
5	.36	.56	.68	.77	.83	.88	.92	.95	.98
6	.40	.60	.72	.80	.86	.90	.93	.96	.98
7	.44	.64	.75	.82	.88	.91	.94	.97	.98
8	.47	.67	.77	.84	.89	.92	.95	.97	.99
9	.50	.69	.79	.86	.90	.93	.96	.97	.99
10	.53	.71	.81	.87	.91	.94	.96	.98	.99

Note. Table entries are Cronbach’s alpha for standardized items.

Figure 1: Composite Factor Model