2013

Predictive Validity and Formative Measurement in Structural Equation Modeling: Embracing Practical Relevance

Jan-Michael Becker
University of Cologne, j.becker@wiso.uni-koeln.de

Arun Rai
Georgia State University, arunrai@gsu.edu

Edward Rigdon
Georgia State University, erigdon@gsu.edu

Recommended Citation

PREDICTIVE VALIDITY AND FORMATIVE MEASUREMENT IN STRUCTURAL EQUATION MODELING: EMBRACING PRACTICAL RELEVANCE

Completed Research Paper

Jan-Michael Becker
Department of Marketing and Brand Management, University of Cologne, Cologne, 50923, GERMANY
j.becker@wiso.uni-koeln.de

Arun Rai
Center for Process Innovation and Department of Computer Information Systems, Robinson College of Business, Georgia State University, Atlanta, GA 30303, U.S.A.
arunrai@gsu.edu

Edward Rigdon
Department of Marketing, Robinson College of Business
Georgia State University, Atlanta, GA 30303, U.S.A.
erigdon@gsu.edu

Abstract

Composite-based methods like partial least squares (PLS) path modeling have an advantage over factor-based methods (like CB-SEM) because they yield determinate predictions, while factor-based methods' prediction is constrained in this regard by factor indeterminacy. To maximize practical relevance, research findings should extend beyond the study's own data. We explain how PLS practices, deriving, at least in part, from attempts to mimic factor-based methods, have hamstrung the potential of PLS. In particular, PLS research has focused on parameter recovery and overlooked predictive validity. We demonstrate some implications of considering predictive abilities as a complement to parameter recovery of PLS by reconsidering the institutionalized practice of mapping formative measurement to Mode B estimation of outer relations. Extensive simulations confirm that Mode A estimation performs better when sample size is moderate and indicators are collinear while Mode B estimation performs better when sample size is very large or true predictability (R^2) is high.

Keywords: Predictive Validity, Formative Measurement, Structural Equation Modeling, PLS path modeling, Factor Indeterminacy

Introduction

Partial least squares path modeling (PLS path modeling; Wold 1982; Lohmöller 1989) and covariance-based structural equation modeling (CB-SEM; Jöreskog 1978, 1982) stand as exemplars of two families of statistical methods—one composite-based and one factor-based—for studying relationships among conceptual variables that are not directly observable (hereafter, conceptual variables). While both approaches have been applied by IS scholars, the greater proportion of articles in the leading IS journals has used PLS path modeling than CB-SEM (Ringle et al. 2012).

At one level, "everyone knows" that composite-based PLS path modeling and factor-based CB-SEM are fundamentally different. It is all there in print, from the earliest discussions of PLS path modeling to recent reminders in the form of prescriptive guidelines on when PLS path modeling or CB-SEM should be applied (e.g., Gefen et al. 2011). And yet, IS researchers (and those in other disciplines) can easily be—and have been—drawn into thinking that the two methods do precisely the same thing, though in different ways: For instance, a researcher may use CB-SEM or PLS path modeling to evaluate models and hypotheses involving conceptual variables which are abstract and directly unobservable. Indeed, a quick look at the IS literature reveals that notwithstanding the guidance on when to apply PLS path modeling and CB-SEM, IS researchers have used—and continue to use—these two techniques in *very similar manners* to evaluate the quality of measurement models and test hypothesized structural effects. This problematic institutionalized practice in the IS discipline has turned the debate about the comparative advantages between PLS path modeling and CB-SEM to a horse race. The focus of this work has been to determine the "winning technique" based on bias in parameter recovery in factor models (e.g., Goodhue et al. 2007, 2012; Reinartz et al. 2009). Given this evaluation criterion to compare the techniques, a recent article has even suggested that PLS path modeling's use should cease (Rönkkö and Evermann 2013). Our contention is that this state of affairs of PLS path modeling vs. CB-SEM has emerged due to an unfortunate blurring of the *fundamental* distinctions between factor-based and composite-based approaches, which has hamstrung the distinctive potential of each of these techniques in furthering IS scholarship and more broadly research in the behavioral sciences. Specifically, while the IS field has herded towards comparing CB-SEM and PLS path modeling based on a *singular* focus on parameter recovery, it has overlooked another key criterion—predictive validity (Shmueli and Koppius 2011).

We recommend reframing the methodological discourse on the comparative advantages of PLS path modeling and CB-SEM from benchmarking the two techniques on one criterion (i.e., parameter recovery) to evaluating how a given technique is used effectively given its objectives and a suitable set of corresponding evaluation criteria (e.g., when prediction can be included as a criterion).

As a factor-based method, CB-SEM defines common factors to represent conceptual variables, while PLS path modeling defines weighted composites for that purpose. Parameter estimates in factor models are chosen to minimize overall discrepancy between empirical and model-implied moments subject to model constraints, while weights in PLS path modeling are chosen to minimize squared error in specific predictive relationships. The use of factors to empirically represent conceptual variables means that the researcher has the opportunity to evaluate the results against a fully specified and highly constrained ideal case, which enables a statistical test of fit. Conforming strictly to this ideal yields advantages in factor model parameter estimation: factor-based approaches can be expected to outperform composite methods in accurately estimating the parameters of correctly specified factor model (Goodhue et al. 2012).

Within the factor model, the representation of a conceptual variable as a common factor is sufficient and even has advantages, but for other purposes such as prediction the common factor representation creates problems. On the PLS path modeling side, conceptual variables are represented by weighted sums of observed variables, using only those observed variables that are specifically associated with each conceptual variable. PLS path modeling yields a vector of weights which, multiplied by the values of the observed variables, produces a specific determinate score for each case in the sample. These scores will have relationships with other variables, within or outside a given model, which are predictable given knowledge of the pattern of covariances of the original observed variables with those other variables. These obvious-seeming results do not apply to factor representations of conceptual variables, due to the often poorly understood phenomenon of factor indeterminacy (Guttman 1955; Maraun 1996; Steiger 1979a). Our **first objective** is to elaborate the differences between composite-based methods and factor-based methods and the implications of those differences for the overlooked evaluation criterion of

prediction.

Following a discussion of the advantages of PLS over CB-SEM for prediction and CB-SEM over PLS for parameter recovery, we will focus on formative measurement in PLS as formative measures are especially suitable for composite-based modeling (Bagozzi 2011) and thus for prediction. At the onset it is clear that the composite-based approach in PLS path modeling has the advantage of easily and intuitively handling formative measures, because composites are weighted linear combinations of the observed variables. Building such a weighted composite of the observed variables naturally fits the formative logic of measurement (e.g., Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003). However, composites are a special case of a formative measurement model with the assumption that the composite is completely represented by the observed (formative) indicators and thus does not incorporate a "construct error term". While some people might view such a composite as a restriction, it provides greater flexibility when formative measures are involved in a model because it does not impose strong constrains for identification (i.e., a composite in PLS path modeling only needs one adjacent composite, either exogenous or endogenous, for identification). Compared to factor-based SEM this composite formation in PLS path modeling has the advantage that it facilitates the use of formative measures without needing the 'two or more dependents' identification rule[1] (Bollen and Davies 2009; Diamantopoulos 2011). Moreover, because of the explicit calculation of composite scores and the sequential (rather than simultaneous) estimation of inner weights (linking different composites) and outer weights (linking each composite to its associated observed variables), PLS path modeling allows for the incorporation of formative composites in endogenous positions to other constructs which can be problematic in factor-based SEM (Cadogan and Lee 2013; Rigdon 2012b). In factor-based SEM, a factor is defined <u>not</u> by the variables that predict it, but by the shared variance of the variables that are dependent upon it (Lee et al. 2013; Rigdon 2013a) and all predictors of a factor are collected into a single equation. If the factor is measured formatively, so that it is dependent on observed variables and the factor is endogenous, so that it is also dependent on other factors, both sets of predictors (observed variables and factors) appear in the same equation. Thus, they "compete" in explaining the shared variance of the variables that depend on the factor. However, this is generally not what the modeler had in mind when specifying the model with a formatively measured construct. In contrast, in composite-based PLS path modeling, the structural predictors and the observed variable predictors appear in two separate equations, each of which has a specification that is consistent with the modeler's intent of specifying a formatively measured construct in a position where it is dependent on other constructs and formed by the observed variables.

The PLS literature offers two main modes for estimating composite scores, which have long been known as Mode A and Mode B. The PLS literature has long suggested that the weights of composites for reflective measures are to be estimated using Mode A while the weights of composites for formative measures are to be estimated using Mode B. This mapping of type of measure (i.e., reflective/formative) to mode of estimation (i.e., Mode A/Mode B) dates back to Fornell and Bookstein (1982) and is repeated in many tutorial articles about PLS (e.g., Chin 1998, Henseler et al. 2009, Hair et al. 2011). This proposed mapping has not received much scrutiny, and it has led researchers using PLS path modeling to mechanically equate formative measurement with Mode B estimation and reflective measurement with Mode A estimation. We challenge this core assumption-in-practice. In reality, of course, PLS path modeling remains a composite-based method, and both Mode A and Mode B estimation create composites of observed variables and therefore are arguably possible choices in how formative measures are specified and composites are computed in a given model. Our **second objective** is to untether the use of formative measurement and the estimation of the composite weights in Mode B. We will show that formative measurement in PLS path modeling does not necessarily need to be based on Mode B estimation and, to the contrary, there are advantages in using Mode A estimation for formative measurement in certain commonly occurring conditions.

Following Dijkstra (2010) and Rigdon (2013b), it is more natural to link Mode B to what are known as regression weights and to link Mode A to what are known as correlation weights. Regression weights, the standard in ordinary least squares regression analysis, take account of both (a) the correlation between each predictor and the dependent variable and (b) the intercorrelations among the predictors. Correlation weights, by contrast, ignore correlations among the predictors. The distinction between correlation

---

[1] If the formatively measured construct is specified without an error term in factor-based SEM, the model needs only one endogenous latent variable to be identified (Diamantopoulos 2011).

weights and regression weights is a standard topic in the forecasting literature (e.g., Dana and Dawes 2004, Waller 2008, Waller and Jones 2010). Building on this literature, our **third objective** is to investigate when correlation weights or regression weights (Mode A or Mode B) should be used to estimate the weights of a formative composite under given conditions. For this purpose we will use a simulation model to assess the properties of both modes under a broad range of conditions.

Simulation studies on PLS path modeling (including simulations in IS) have predominantly focused on parameter recovery and Type I/Type II errors. Moreover, they have focused on factor models and comparison to factor-based methods (e.g., CB-SEM). Consistent with the first and second research objectives, our simulations will move beyond factor models and utilize a completely formative composite population model. In addition, the evaluation criteria for this study will include not only parameter recovery but also predictive ability. This will demonstrate how investigating the predictive abilities of PLS path modeling can yield additional insights in the evaluation of a model of unobserved conceptual variables.

Our results will help IS and other social science researchers to better understand the differences between factor-based and composite-based approaches and the usefulness of the latter to facilitate prediction. Moreover, the results of the simulation study will offer guidelines to researchers using PLS path models with formative composites in their choice of the best approach to estimate the weights of the composites and thus derive more reliable conclusions.

Untethering PLS from Factor-based Methods

From its origin, PLS path modeling has been tied closely to factor-based CB-SEM. Herman Wold created PLS path modeling under the inspiration of the advances made by his student, Karl Jöreskog, in factor-based CB-SEM (Wold 1988). While Wold originally presented his "basic design" in terms of a "second order factor model with single indicators" (Dijkstra 2010, p. 24), using factor model language to describe his new approach, Wold also emphasized the distinctions between the methods, seeing in those distinctions the real merit of PLS path modeling, not as a competitor to factor-based SEM but rather as a complement (Wold 1982). Today, most sources make the distinction clear, identifying PLS path modeling as a multiple indicator modeling technique based on composites (Esposito Vinzi et al. 2010, Rigdon 2013b).

Moreover, the close connection between psychological measurement and factor analysis might lead a researcher to mistake PLS path modeling for a factor-based method. Both common factor analysis and the "classical" true score theory of measurement were introduced to the academic literature by Charles Spearman, in separate papers both published in 1904 (McDonald 1999). Indeed, true score theory is merely a special case of the common factor model (Steiger and Schönemann 1978). Highly influential treatments of scale development by Churchill (1979) and Gerbing and Anderson (1988) point to factor analysis and no other technique as the path to developing sound measures. Scholars could be forgiven for believing that "measurement" and "factor analysis" are synonymous, or that any non-factor-based method must have an inherent and substantial weakness when it comes to measurement.

Yet, factor-based methods also have profound weaknesses. Among those, factor models are plagued by factor indeterminacy (Guttman 1955, Maraun 1996, Steiger 1979a). A covariance matrix of p observed variables has a rank of at most p, meaning that there are at most p dimensions of information contained within the data (Mulaik 2010). A factor model for a set of p observed variables, all congeneric indicators of one common factor, typically involves p unobservable error terms—one for each observed variable—plus the common factor. Basic assumptions specify that the error terms are all distinct, and often mutually orthogonal, and that the common factor is orthogonal to all p error terms. Thus, the joint covariance matrix of the error terms and the common factor must have rank p + 1, *higher* than the rank of the data from which the model is estimated (Steiger 1979b). This statistical trick is achieved by making the error terms and common factor indeterminate. This means that one cannot assign one best value to the common factor and the error terms for each case or observation in the dataset. In factor analysis there are an infinite number of different sets of "factor scores" that may be assigned to the cases in a dataset. Each of these sets of scores is perfectly consistent with the specified factor model and adheres to all constraints in the model, yet the various sets need not be very highly correlated with each other. For a given factor model—even if all assumptions hold—a given case within the data set might be assigned any of a range of scores to represent the common factor in the model.

One crucial consequence of this indeterminacy is that the correlation between a factor and any variable outside the factor model is itself indeterminate (Schönemann and Steiger 1978)—it may be high or low, depending on which set of factor scores one chooses. The range of this correlation is a function of both the correlation between the factor model's observed variable indicators with the external variable and the indeterminacy of the factor (Steiger 1979b). The correlation between factor and outside criterion may be high even if the observed variables loading on the factor are orthogonal to the outside criterion (Steiger 1979b). In fact, for any factor model and any outside criterion variable, there is some set of scores for the common factor and the specific factors / error terms that together will predict the criterion perfectly (Schönemann and Steiger 1978), even if the criterion is pure white noise (Schönemann and Haagen 1987). Hence, factor models are *grossly unsuitable* for prediction.

Most factor modeling software will generate what are called "estimated factor scores" or "latent variable scores," but these are not true factor scores at all, but rather geometric averages across the ranges of possible factor scores (Mulaik 2010). These ersatz scores will, indeed, reproduce the parameters of the factor model from which they are drawn, but they will not conform to all of the constraints which are part of the factor model (Steiger 1979b). This should be enough to make the use of such scores in research at least suspect.

In contrast, methods of weighted composites, including PLS path modeling, have their advantages as they immediately produce a single specific value for each composite for each case, once the weights are established. Outside of artificial situations where factor model assumptions hold perfectly and true parameter values are known, composite methods may yield parameter estimates that are substantively indistinguishable from the estimates of factor models for the same data (Velicer and Jackson 1990). And composite methods have the advantage of yielding determinate scores for each composite for each case or respondent, which can be used beyond the specific model in which the scores were created. These determinate scores are not factor scores and certainly not the indeterminate factor but rather proxies of the concepts under investigation in PLS path modeling, just as factors are proxies for the conceptual variables in factor-based SEM.

PLS Evaluation Criteria Must Include Predictive Validity

In order for scientific research to have value, its findings must be validated—demonstrated to be sound, within the limits of our ability to do so. Classical treatments of validation methods, such as Messick's (1989) seminal chapter or Shadish, Cook and Campbell's (2002) authoritative volume on experiment design, describe many types of evidence that can be marshaled, under headings including "content validity," "construct validity," and "predictive validity." Yet, somehow the structural equation modeling literature seems to have fixated entirely on "construct validity," particular to the exclusion of predictive validity. Perhaps this neglect of predictive validity is a consequence of the weakness of indeterminate factor models in predictive applications. Regardless, researchers cannot afford to ignore predictive validity. Nobel laureate in economics Milton Friedman (1953, pp. 8-9) declared, "… the only relevant test of the validity of a hypothesis is comparison of its predictions with experience."

In most any study, the larger part of cases of interest lie outside the sample of data used to estimate a given model. To expand knowledge, researchers look beyond any given set of phenomena for additional connections and for new opportunities to challenge their understanding. Using data and phenomena that lie outside a given model, predictive validity provides a truly objective standard for evaluating success. Predictive validity also provides a criterion which practitioner audiences immediately grasp. An emphasis on predictive validity might help to combat the common problem of unreplicable findings in the social sciences (Yong 2012). When all criteria for evaluating results lie within the model or study itself, the potential for researchers to influence those criteria is large. Emphasizing predictive validity will mean that at least some of these criteria are beyond researchers' reach. Researchers who ignore predictive validity are not applying the most rigorous methods to evaluate their models and theories.

In general, multiple indicator approaches should have an advantage in prediction over single indicator methods. Within the literature on forecasting, it is well established that combinations of forecasts outperform single forecasts (Makridakis and Hibon 2000; Makridakis and Winkler 1983). A combination of forecasts can be more accurate than the most accurate single forecast in the combination (Armstrong 2001a; see also Diamantopoulos et al. 2012). Employing multiple indicators for conceptual variables should similarly improve prediction. Better forecasting results also follow from reducing a criterion to be

forecast to its components and then forecasting each component individually, rather than making a single forecast for the variable overall (Armstrong 2001b). Formative measurement of proxies for the conceptual variables in a PLS path model is certainly consistent with this "best practice" from the forecasting literature.

Formative Measurement in PLS Need Not be Limited to Mode B

By "formative measurement," researchers typically mean representing some conceptual variable by forming a weighted composite of observed variables thought to be associated with that conceptual variable. PLS path modeling users readily agree that PLS Mode B estimation equates to formative measurement, because, for each conceptual variable, the method regresses a composite on a set of observed variable predictors. PLS Mode A estimation does precisely the same thing, though somewhat differently. In consequence, there is no reason to limit estimation of "formative measurement models" to Mode B, and there is good reason to consider estimating such models using Mode A.

Since PLS path modeling is a composite-based method, it should follow that even when Mode A is used, the estimation still involves a composite and a set of observed variables. The difference is that, while Mode B does this regression using all observed variable predictors as a set, thus taking account of collinearity among the predictors, Mode A performs this regression one predictor at a time, thus ignoring collinearity among the predictors (Dijkstra 2010, Rigdon 2012b). Using formulas derived by Dijkstra (2010, pp. 32-33), Rigdon (2012b) argued that the distinction between Mode B and Mode A paralleled the distinction between "regression weights," which take account of predictor collinearity, and "correlation weights," which ignore collinearity (Waller and Jones 2010).

Regression weights adjust for collinearity so as to give less weight to more redundant predictors. This enables the optimal properties of OLS regression weights when assumptions hold. In particular, regression weights optimize R^2 for the data which are used to estimate the weights (in-sample prediction). The performance advantages of "optimal" regression weights are not necessarily overwhelming or universal. In many cases, substantially different weights may perform almost as well as "optimal" regression weights (Waller 2008). There are also a range of situations where other types of weights, including correlation weights (Dana and Dawes 2004), simple unit or equal weights (Einhorn and Hogarth 1975), or even totally random weights (Wainer 1976) perform predictably *better* than "optimal" regression weights. Dana and Dawes (2004) demonstrated, in the context of conventional regression, that correlation weights outperform regression weights particularly when it comes to out-of-sample prediction, where weights are first estimated in one dataset and then used to predict a criterion from a different dataset.

When researchers assess predictive validity, they are evaluating out-of-sample predictive ability. Dana and Dawes (2004) found that regression weights were only superior under very specific conditions. They urged researchers to avoid using regression weights for out-of-sample prediction. It is plausible that PLS path modeling under different modes will support similar conclusions, giving the predictive validity edge to Mode A under a broad range of conditions. However, no study has ever examined this question empirically.

Overview of PLS Simulation Studies Guiding IS Research

Many simulation studies in the IS literature and other fields have guided the use of PLS path modeling and CB-SEM in IS research. We reviewed simulation studies that investigate PLS' behavior under various model constellations. We include in our review the information systems literature, in particular the field's two widely recognized premier journals (*MIS Quarterly* and *Information Systems Research*), recent (2012) papers from the *International Conference on Information Systems* as well as recent special issues on PLS path modeling (*Long Range Planning* and *European Journal of Information Systems*) and marketing. Based on our review, we identified 14 simulation studies that reported results about the behavior of PLS in simulation studies (Table 1), and make three observations:

1. Almost all these studies use a factor-based population model (i.e., reflective measures). Only some very recent exceptions (e.g., Aguirre-Urreta and Marakas (2012); Becker et al. (2013); Henseler et al. (2012)) use formative measures in exogenous positions in a PLS path model. We did not identify any simulation study that involves a path model with formative measures in an endogenous position. This is surprising as one of the major advantages of the composite-based PLS method is its flexibility in the use of formative composites compared to CB-SEM which imposes several constrains regarding the use of formative

measures.

2. All the simulation studies focus on parameter recovery and statistical power, with the assessment of predictive validity rarely considered. Only one recent study (i.e., Evermann and Tate 2012) investigates predictive abilities of PLS path modeling in terms of the blindfolded cross-validation Q^2 and in-sample R^2. However, the assessment of the Q^2 is only applicable to reflective measures (factor models) as it predicts missing values in observed data and is therefore not suitable for formative composites. Moreover, Evermann and Tate (2012) used a model which exhibited high predictive performance (i.e., population R^2 of 0.9). This limited consideration of predictive validity in evaluating PLS path modeling is surprising as a noted key advantage of PLS path modeling is its predictive ability.

3. Given the focus on parameter recovery and statistical power, most simulation studies do not investigate the in-sample as well as out-of-sample predictive abilities (e.g., the R^2 value). In fact, only four studies investigate the squared correlation between true and estimated composite scores. While two of them refer to it as prediction accuracy (Henseler and Chin 2010; Henseler et al. 2012), the other two refer to it as composite reliability (Becker et al. 2012; Rönkkö and Evermann 2013)[2]. Nevertheless, it assesses only in-sample properties and does not investigate prediction beyond the sample characteristics. We are not aware of any study that investigates out-of-sample prediction in PLS simulation studies.

In contrast, our simulation setup will evaluate the performance of the two alternative modes of computing composites in PLS path models (i.e., Mode A or Mode B) by (1) using a model that includes formative composites not only in exogenous but also in endogenous position, (2) focusing on parameter recovery *and* prediction, (3) investigating the in-sample as well as out-of-sample predictive abilities, and (4) using a true composite-based population model.

Table 1. Summary of Key PLS Simulation Studies

Authors Year Journal	Objective of Simulation Study	Model	Evaluation Criteria
Chin et al. 2003 *Information Systems Research*	Investigate interaction-effects in models with measurement error; compare PLS to single-indicator regression and summated regression	reflective / factor-model	parameter recovery (mean estimates; mean relative bias) and statistical power (type II errors)
Goodhue et al. 2007 *Information Systems Research*	Comparing PLS and Regression in detecting interaction effects	reflective / factor-model	parameter recovery (mean estimates; mean relative bias) and statistical significance and power (type I and II errors)
Qureshi and Compeau 2009 *MIS Quarterly*	Compare PLS and CB-SEM multi-group methods	reflective / factor-model	statistical power (type II errors) to detect between group differences
Reinartz et al. 2009 *International Journal of Research in Marketing*	Compare PLS and CB-SEM	reflective / factor-model	parameter recovery (relative bias, mean absolute relative bias) and statistical power (type II errors)
Henseler and Chin 2010 *SEM Journal*	Investigate different methods to model interaction effects in PLS	reflective / factor-model	parameter recovery (mean relative bias), statistical power (type II errors) and prediction accuracy (squared correlation between estimated and true composite scores)

---

[2] The term composite reliability for the squared correlation between true and estimated composite score is also used by other authors outside of the PLS context, e.g., Bollen and Lenox (1991).

Hwang et al. 2010 *Journal of Marketing Research*	Comparing GSCA, CB-SEM and PLS	reflective / factor-model	parameter recovery (mean absolute bias)
Lu et al. 2011 *International Journal of Research in Marketing*	Comparing two "new" CB-SEM estimators to ML CB-SEM and PLS	reflective / factor-model	parameter recovery (relative bias; mean absolute bias), coverage (confidence intervals) and statistical power (type II errors)
Becker et al. 2012 *Long Range Planning*	Comparing different methods to model formative hierarchical components in PLS	formative hierarchical component / reflective (all others)	parameter recovery (root mean squared error) and composite score reliability (squared correlation between estimated and true composite scores)
Chin et al. 2012 *MIS Quarterly*	Investigate PLS methods to address common method bias	reflective / factor-model	parameter recovery (mean estimates (&standard deviation of estimates))
Goodhue et al. 2012 *MIS Quarterly*	Compare PLS, CB-SEM and Regression	reflective / factor-model	parameter recovery (mean estimates (&standard deviation of estimates); mean relative bias) and statistical significance and power (type I and II errors)
Henseler et al. 2012 *European Journal of Information Systems*	Investigate different methods to model quadratic-effects in PLS	formative (exogenous) – reflective (endogenous)	parameter recovery (mean relative bias), statistical significance and power (type I and II errors) and prediction accuracy (squared correlation between estimated and true composite scores)
Becker et al. 2013 *MIS Quarterly*	Comparing different segmentation methods in PLS	formative (exogenous) – reflective (endogenous)	parameter recovery (mean absolute bias) of group estimates
Rönkkö and Evermann 2013 *Organizational Research Methods*	Examination of common beliefs about PLS path modeling	reflective / factor-model	parameter recovery (bias), significance and power (type I and II errors) and composite score reliability (squared correlation between estimated and true composite scores)
Aguirre-Urreta and Marakas 2012 *ICIS*	Comparing PLS and CB-SEM when omitting formative indicators	formative (exogenous) – reflective (endogenous)	parameter recovery (mean estimates; mean percentage bias)
Evermann and Tate 2012 *ICIS*	Comparing the predictive abilities of PLS and CB-SEM	reflective / factor-model	predictive abilities (mean communality-based and redundancy-based Q^2; mean R^2)
Sharma and Kim 2012 *ICIS*	Model Selection in PLS	reflective / factor-model	model selection criteria

Simulations

We conducted extensive simulations to evaluate the use of Mode A (correlation weights) and Mode B (regression weights) for formative measures in PLS path models, with the aim of evaluating the predictive abilities of the modes. Based on this objective, we employed a simulation design similar to that of Dana

and Dawes (2004).

Simulation Design
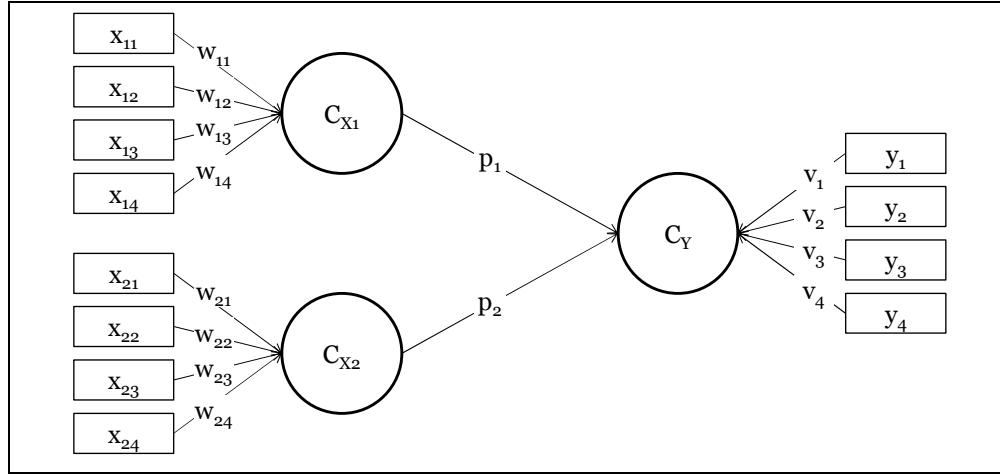
Figure 1. Population Path Model

Figure 1 depicts our population path model. Our simulation study was based on a population composite model (not a factor-model) containing two exogenous composites, C_{X1} and C_{X2}, and one endogenous composite, C_Y, via the single structural equation:

$$C_Y = p_1 * C_{X1} + p_2 * C_{X2} + e$$

In this simulation, each composite has four observed variable components so that each composite is formed as:

$$C_{X1} = w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + w_{14}x_4 = \sum_{i=1}^{4} w_{1i}x_i$$

$$C_{X2} = w_{21}x_5 + w_{22}x_6 + w_{23}x_7 + w_{24}x_8 = \sum_{i=5}^{8} w_{2i}x_i$$

$$C_Y = v_1 y_1 + v_2 y_2 + v_3 y_3 + v_4 y_4 = \sum_{i=1}^{4} v_i y_i$$

It will be convenient to restate these relations in matrix terms. Let $C_{X\,(2x1)}$ be a column vector of the two exogenous composites, let $X_{(8x1)}$ be a column vector of the observed variable components of the C_X, and let $Y_{(4x1)}$ be a column vector of the components of Y. Define two component weight matrices $W_{(2x8)}$ and $V_{(1x4)}$. Then we can concisely restate relations between composites and components as:

$$C_X = WX$$

$$C_Y = VY$$

In the population, these equalities hold exactly, without an error term. To define the population, we specified a population correlation matrix, $\Sigma_{(12x12)}$, which can be represented as a partitioned matrix:

$$\Sigma = \begin{matrix} \Sigma_{xx} & \Sigma_{xy}^T \\ \Sigma_{xy} & \Sigma_{yy} \end{matrix}$$

where $\Sigma_{xx(8x8)}$ contains the population correlations among the components of the two exogenous composites, $\Sigma_{yy(4x4)}$ similarly for the endogenous composite, and $\Sigma_{xy(8x4)}$ contains the cross-correlations for the components of the exogenous and endogenous composites.

If this were a factor model, then correlations among the indicators of a given factor would be determined by factor loadings. Here we are working with composites, so we must specify the pattern of correlations among the components of the exogenous indicators and among the components of the endogenous composite. In practice, different components of a given composite may be correlated or uncorrelated, so this simulation used a pattern of population correlations that covers a range of possibilities. We defined the correlation matrix of the components of the endogenous composite as:

$$\Sigma_{yy}=R_y \qquad R_y=\begin{matrix} 1 & k & 0 & 0 \\ k & 1 & 0 & 0 \\ 0 & 0 & 1 & k \\ 0 & 0 & k & 1 \end{matrix}$$

Thus, in the set of four components, we have two pairs where the members of each pair are correlated among themselves but are uncorrelated with the members of the other pair. We imposed the same pattern upon the correlations of the components of the exogenous composites:

$$\Sigma_{xx}=\begin{matrix} R_x & 0 \\ 0 & R_x \end{matrix} \qquad R_x=\begin{matrix} 1 & k & 0 & 0 \\ k & 1 & 0 & 0 \\ 0 & 0 & 1 & k \\ 0 & 0 & k & 1 \end{matrix}$$

The last remaining task is to calculate Σ_{xy}. In calculating the cross-correlations we have to account for PLS path modeling's routine standardization of composites. It is well known that the variance of a weighted composite σ^2 is equal to a weighted sum of the component variances and two times a weighted sum of the component covariances. Even if the observed variable components are themselves standardized, it is unlikely that the resulting composite will have unit variance. Therefore, we standardize the weights of the composites with the standard deviations σ_{CX} and σ_{CY} so that the resulting composite has unit variance. In addition, this makes the weights comparable to the estimated weights in PLS path modeling:

$$W^* = W / \sigma_{CX} \qquad V^* = V / \sigma_{CY}$$

Recall that the structural model included two slope coefficients p_1 and p_2, which can be collected in a row vector P. In designing the simulation, we set standardized parameter values p_1 and p_2, so that we can control the population R^2 value for the dependent composite.

Finally, we can calculate the cross-correlations between the components of the composites Σ_{xy}:[3]

$$\Sigma_{xy}=\Sigma_{yy}V^{*T}PW^*\Sigma_{xx}$$

To verify that this population correlation matrix complies with the intended properties, one can use the true standardized weights W^* and V^* and calculate the composites C_X and C_Y on population data which will exhibit the pre-specified correlations between the composites. Alternatively, one can use standard formulas to calculate the correlations between composites of components and get the same result:

$$Z^T\Sigma Z = \begin{matrix} 1 & 0 & p_1 \\ 0 & 1 & p_2 \\ p_1 & p_2 & 1 \end{matrix} \qquad with \qquad Z=\begin{matrix} W & 0 \\ 0 & V \end{matrix}$$

As population values, we used the same set of four <u>unstandardized</u> component weights for all three composites: $w_{j1} = 0.7$, $w_{j2} = 0.6$, $w_{j3} = 0.3$, and $w_{j4} = 0.25$. The path coefficient between C_{X1} and C_Y, p_1, is always 0.3 and the path coefficients between C_{X2} and C_Y is defined to be $p_2 = sqrt(R^2 - p_1^2))$, with R^2 being one of our varying simulation design parameters.

We vary three design-factors in our simulation study:

    i.   The *R^2* (the true predictability) of the endogenous composite by the two exogenous composites: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8 (thus covering a broad range).

---

[3] Note: These cross-correlations are different from a factor-model which would be calculated as $\Sigma_{xy}=L_y^TPL_x$ with L being a loadings vector.

    ii.   *The correlation k between pairs of components* for each composite. We increase the correlation k of each pair of components from 0 to .95 in steps of .05 (again covering a broad range).

    iii.   The *Sample Size*: 40, 100, and 500 (reflecting levels encountered in IS applications).

After defining the population model covariance matrix for each simulation condition, we generated a population dataset from the population model with a very large sample size (10,000 observations) with random variables from the multivariate normal distribution. Each of these populations was used only once. We then draw 40, 100 or 500 observations from the population dataset as the estimation sample.[4] Next, we estimate a PLS path model on the estimation sample specified with all composites in Mode A and again with all composites in Mode B. As a result, we derive the sample estimates for the weights (ws_{11} to ws_{24} and vs_1 to vs_4) and the path coefficients (ps_1 and ps_2) as well as the in-sample R^2 value. Next, we calculate the out-of-sample R^2 by 1) calculating new exogenous composites C_{XS1} and C_{XS2} based on the estimated <u>sample</u> weights (ws_{11} to ws_{24}) for both modes and the overall <u>population</u> dataset, 2) calculating an endogenous composite C_{YS} based on the estimated <u>sample</u> path coefficients and the C_{XS1} and C_{XS2}, calculated in the prior step, and 3) calculating residuals E from the difference between this new composite C_{YS} and the true composite C_Y (i.e., a linear combination of the indicators y_1 to y_4 and the true population weights v_1 to v_4). The out-of-sample R^2 is then: $1 - (E^T E)/(C_Y^T C_Y)$. Moreover, we calculate the estimation error of the weights and path coefficients (root mean squared error (RMSE)) based on the differences between the sample estimates and the true population values.

Simulation Results

We have simulated 1,000 datasets per design-factor combination resulting in 7*20*3*1,000 = 420,000 estimations of Mode A and Mode B results. In addition, we also obtained results for a model where the unstandardized weights (W or V) for each composite were fixed at 1—i.e., unit weights. This was meant to provide a useful standard for comparison.

We will discuss the results of the simulation based on the following three criteria: 1) the squared estimation error (RMSE[5]) of the parameters (i.e., weights and path coefficients), 2) the in-sample prediction (R^2), and 3) the out-of-sample prediction (R^2). Finally, we will draw conclusions based on the evaluation of these three different criteria.

Figure 2 shows the path coefficients' RMSE aggregated over all different R^2 levels in the simulation. We find that the path coefficients' RMSE in very low sample sizes (40 observations) is very similar across the three approaches to form a composite (Figure 2). All three approaches yield rather high RMSE for the two path coefficients. The RMSE decreases when sample size increases, i.e., the parameter recovery gets better. However, the RMSE for unit weights decreases much less than the RMSE for Mode A and Mode B. With larger sample sizes (100 or 500 observations) both Mode A and Mode B do not show a significant difference in the RMSE for the path coefficients. Moreover, analyzing the RMSE for different R^2 levels (Table A.1 in Appendix A), we find that unit weights perform as good as Mode A and Mode B or even better for very small population R^2 and small sample sizes. However, while Mode A and Mode B improve their estimation precision as R^2 and sample size increase, the RMSE of unit weights stays the same or even increases.

In addition, we find that the path coefficients' RMSE does not depend on the multicollinearity of the indicators (Figure 2). However, this should not be surprising as the design permitted multicollinearity only within each set of indicators and not across composites. Therefore, the paths between composites should not be affected by collinearity.

---

[4] We sample without replacement so that each observation is only selected once for the sample dataset, however, observations drawn to the sample dataset are still part of the population dataset to which we calculate the out-of-sample prediction. Dana and Dawes (2004) note that this might favor regression weights (i.e., Mode B). However, with sample sizes much smaller than the population superset, this bias should be very small.

[5] We also assessed the mean absolute error and the mean absolute relative error and did not find any difference in the pattern that we report in the paper. The advantage of absolute or squared measures of error is that they not only include the bias but also the spread (e.g., standard deviation) of the error.

In terms of their ability to recover path coefficients, Mode A and Mode B yield very similar results and thus are equally suited for use with formative composites in PLS path modeling. Unit weights should be avoided when the focus is on the interpretation of the path coefficients unless very small sample sizes are used and low R^2 values are expected. Even then, the estimation error from unit weights will be so high that any interpretations will be questionable.
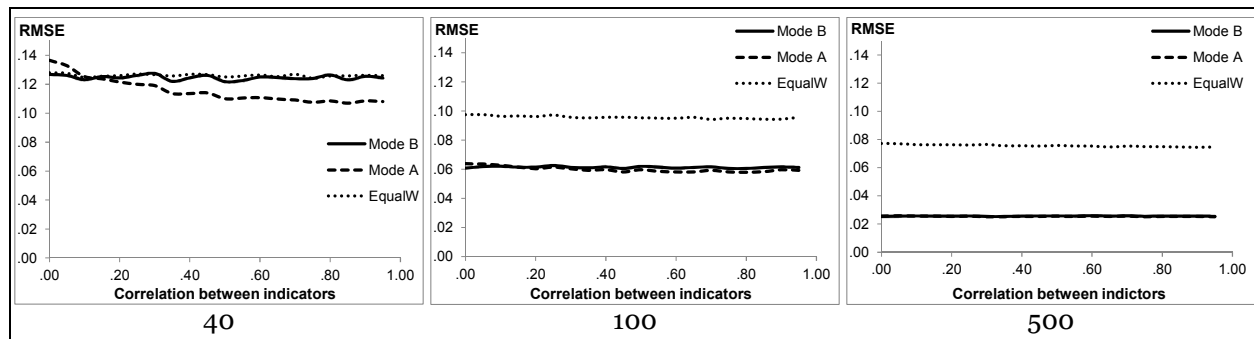

Figure 2. Path Coefficient Estimation Error for Different Sample Sizes

In contrast, the RMSE for the weights depends on the multicollinearity for Mode A and Mode B. Figure 3 shows that the RMSE increases for Mode B when the correlation between the indicators increases. This effect is amplified when sample sizes are low. In contrast, Mode A weights are not as heavily influenced by the correlation between indicators and even produce better results when multicollinearity increases. Investigating this effect more thoroughly, we find that the mean bias of the weights for both modes decreases with higher multicollinearity while the variance of the estimated weights increases for Mode B and decreases for Mode A. Thus, Mode B has a higher RMSE because of the large variance in the estimated weights even though the mean bias decreases, because of sampling variability. The direction of the bias in the weights in all our conditions shows that weights are underestimated in both Mode A and Mode B. This is interesting to note as the PLS literature acknowledges overestimation of loadings (Chin 1998; Rigdon 2013b; Wold 1985) but to the best of our knowledge has not assessed weights.

Unit weights are not affected by sampling variability and only include the squared bias in the RMSE. Estimating the weights with Mode A performs generally better than setting unit weights in terms of RMSE. In contrast, estimating weights with Mode B is only preferable to unit weights when multicollinearity is low and sample sizes are not very small.

When researchers want to interpret the weights of the composites in their research they should use Mode A (correlation weights) unless sample sizes or expected R^2 values are high or multicollinearity is supposed to be very low between the indicators. As the correlation between formative indicators increases Mode B (regression) weights are heavily infected by multicollinearity while correlations weights (Mode A) are not.
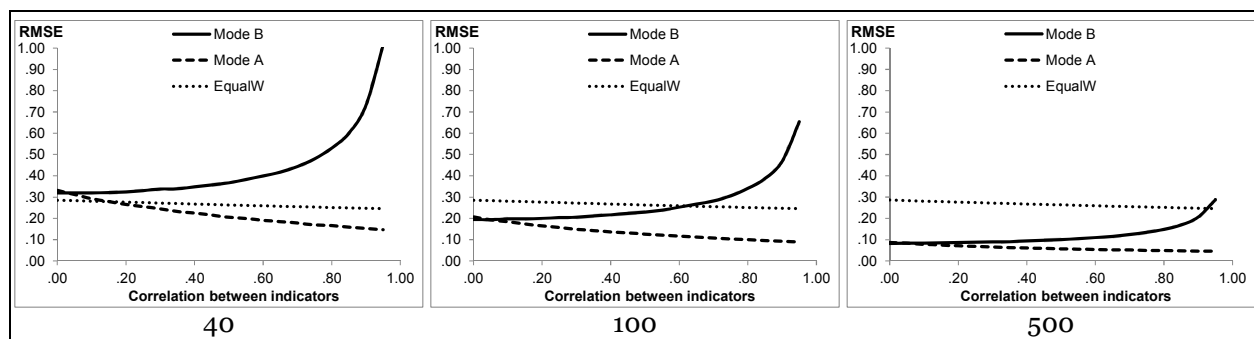
Figure 3. Weights Estimation Error for Different Sample Sizes

In the design of this study, the true predictability of the dependent composite ranges from 0.2 to 0.8, with an overall average of 0.5. Mean overall R^2 values that deviate from 0.5 indicate a bias. PLS path modeling is known for maximizing the in-sample predictive power of the model (Lohmöller 1989). Hence, both Mode A and Mode B show overfitting rates in low sample sizes (Figure 4) or for small R^2 values (Table A.1 in Appendix A). Mode A is overfitting less, i.e., it produces smaller R^2 values that are closer to the true

population R^2 value when sample sizes are small or R^2 is low. As sample size increases and R^2 increases both modes become more similar and overfitting is reduced, showing R^2 values closer to the true population value (Figure 4 and Table A.1). In contrast, unit weights consistently perform worse than both Mode A and Mode B and underestimate the true population R^2 in all conditions.

If the focus of the research is to develop a predictive model that has the highest in-sample predictive power, then Mode B should be employed as it capitalizes most on the characteristics of the dataset, especially when sample sizes are small or R^2 is low. Research that focuses on explanatory models should favor Mode A as it produces sample estimates of the R^2 value that are closer to the population value.
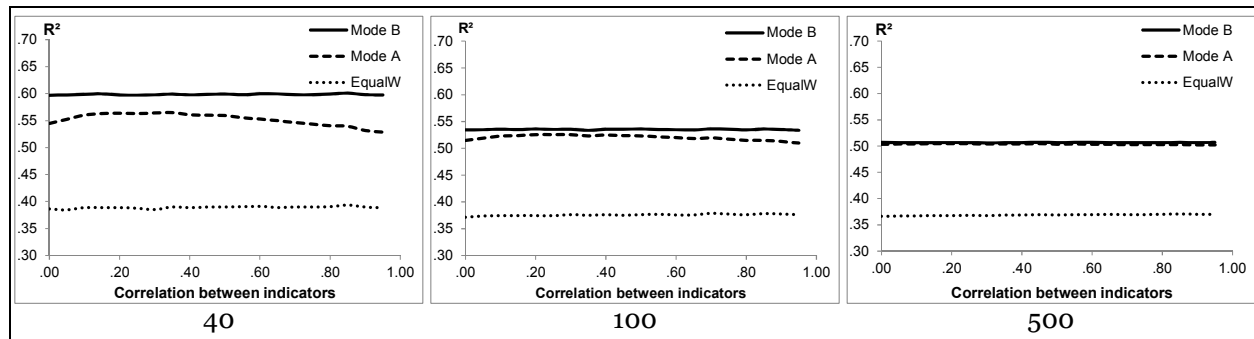
Figure 4. In-Sample R^2 for Different Sample Sizes

Finally, in terms of out-of-sample prediction, we find that unit weights perform very well in small samples and with small population R^2 values (Figure 5). As sample size increases and population R^2 increases Mode A and Mode B perform better than unit weights. Mode A generally performs better than Mode B when sample sizes are not very large and when the R^2 is smaller. Moreover, the out-of-sample prediction of Mode A increases with higher multicollinearity between the indicators of the composites while unit weights and Mode B are unaffected by the collinearity. Mode B only outperforms the other two approaches in large samples where indicators exhibit very small collinearity.

In conclusion, Mode A is the preferable approach for out-of-sample prediction when sample sizes are not large and R^2 is medium to small, especially when indicators are correlated. Unit weights are only preferable in very small samples with low R^2 and small correlations between the indicators.
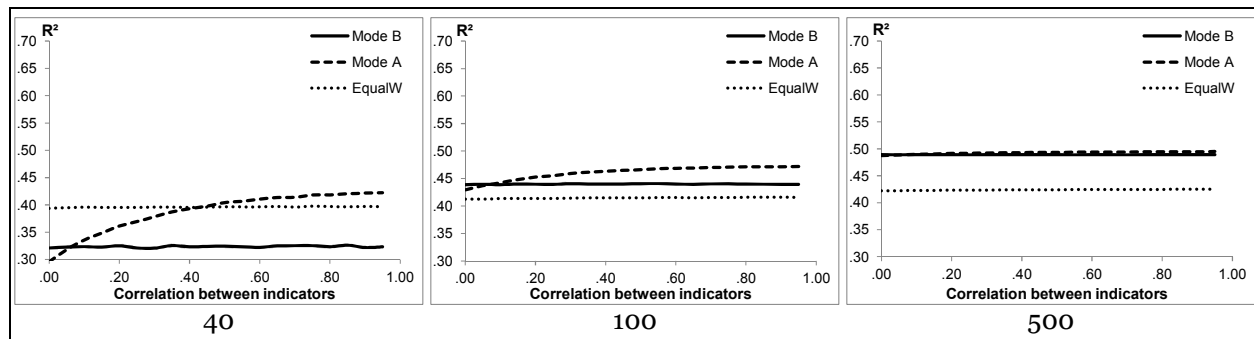
Figure 5. Out-of-Sample R^2 for Different Sample Sizes

The results from the in-sample and out-of-sample prediction analysis are in line with the findings by Dana and Dawes (2004) who investigate correlation and regression weights in a simple regression context. It shows that the usual approach applied in PLS path modeling, i.e., to use Mode B for formative measures, is not always preferable. In fact, our simulations show that in many cases Mode A might be the better approach. Thus, researcher using PLS path models with formative composites should be aware of the different characteristics of correlation and regression weights (for a summary see Table 2) and not revert to regression weights (Mode B) as the only alternative to estimates weights of formative composites in PLS path modeling.

Conclusion

We broaden the discussion of multiple indicator structural equation methods by examining when composite-based approaches are meaningful and how modeling choices should be made on computing composites in PLS path modeling. There has been ample discussion to emphasize that when researchers are working with a pure factor-model and are concerned with testing hypotheses based on this factor-model, they should revert to a factor-based method (e.g., CB-SEM) and not use PLS path modeling (Rönkkö and Evermann 2013; Goodhue et al. 2007, 2012). Yet, composite-based approaches are necessary as there are many situations where researchers find that 1) the model does not comply with the strict assumptions of factor-models (Gefen et al. 2011), 2) measures are composites of observed variables (e.g., conceptual variables measured with formative indicators) (Petter et al. 2007), and 3) the scientific/practical objective is to develop a predictive model (Shmueli and Koppius 2011). In the first case, factor-methods lose their optimality. In the complex world of social science phenomena examined in much IS research, empirical models do not always follow the strict assumptions of factor-models. Hence, non-factor-based methods (e.g., composite-based methods) that do not possess optimality for factor-models, but are more flexible, can perform as well—or even better—than factor-based methods. Actually, *all* previous studies comparing PLS path modeling to CB-SEM have investigated both methods in situations where the assumptions of factor models held. Thus, it is an open question how composite methods, like PLS path modeling, perform in comparison to CB-SEM, when the assumptions of factor-models are not fulfilled. In the second and third case, the use of factor-based methods might not even be possible. Factor-based methods like CB-SEM have restrictions regarding the use of composites in specific positions of the model (e.g., endogenous positions) and posit limiting constrains regarding the identification of the model. In addition, we have discussed why factor-based methods are not usable for prediction of variables or observations outside of the model. All these considerations imply that IS researchers should not limit their methodological choices for modeling conceptual variables with multiple indicators to factor-based methods, but should also consider alternative approaches like composite-based PLS path modeling.

While CB-SEM focuses on optimality, PLS path modeling trades optimality for flexibility. Researchers using PLS will naturally lose some accuracy in parameter recover compared to optimal factor-based methods *if* the model fulfills the restrictive factor-model assumptions. Yet, they also gain something valuable: the ability to predict and the ability to use predictive validity to evaluate the quality of the model. Investigating the model for predictive validity (out-of-sample prediction) allows deriving more generalizable conclusions and can offer complementary insights into phenomena.

Therefore, the assessment of composite-based methods, like PLS path modeling (or regressions based on unit weighted composites), should move away from a sole focus on parameter recovery. It should include predictive validity as a standard assessment when evaluating path models. However, including predictive validity as an important part of model assessment does not only apply to researchers who use PLS path modeling to estimate models with empirical data, but also to methodology researchers who investigate the performance of composite-based methods like PLS path modeling.

We demonstrate the usefulness of predictive analytics in simulation studies by reconsidering the default usage of Mode A for reflective measures and Mode B for formative measures in PLS path modeling. We show that researchers employing formatively measured composites in their models should prefer Mode A over Mode B when they are concerned about predictive validity, except for the situations where sample size is very large or population R^2 is high and multicollinearity is low (Table 2). Moreover, the simulations confirm that out-of-sample predictive validity from PLS analysis will generally be poor unless sample size is at least moderate. This finding is in line with recent studies based on parameter recovery questioning the usefulness of PLS path modeling for small sample sizes (Goodhue et al. 2012; Rönkkö and Evermann 2013). When sample size and population R^2 is low, unit weights provide the best out-of-sample prediction. However, with medium or large sample sizes or large population R^2 Mode A weights outperform unit weights and offer good out-of-sample predictive validity.

Reconsidering the use of Mode B for formative composites might also open up the question of the default use of Mode A when measurement is believed to be reflective. However, our results indicate that Mode A outperforms Mode B in situations with high multicollinearity. In contrast to formative measures, reflective measures should exhibit high correlations among the indicators and thus our results indicate that Mode A should continue to be used in such cases. Nevertheless, the general concerns about low predictive validity

of composites in very small samples are probably also transferable to this situation.

Table 2. Summary of Simulation Conclusions

	Mode A (Correlation Weights)	Mode B (Regression Weights)	Unit Weights
Bias/ Estimation Error	Generally, best for the estimation of path coefficients in small to medium samples or low to medium R² values. Best, for the estimation of weights with correlation between indicators.	Performs as well as Mode A in medium to large samples for the path coefficients and should not be used for estimation of weights when strong collinearity affects the indicators.	Only preferable for the estimation of path coefficients in cases with very low R² and very small samples sizes. Not usable for the <u>estimation</u> of weights.
In-Sample Prediction	Generally offers in-sample R² values that are closest to the population value.	Overestimates the in-sample R² value for small sample sizes and low R² values.	Generally not preferable in terms of in-sample prediction, as the R² is always underestimated.
Out-of-Sample Prediction	Generally preferable over Mode B when strong collinearity exist between the indicators. Preferable over Mode B when sample sizes and R² are both not very small.	Performs as well as Mode A in large samples or high R² situations. Is only preferable over the other approaches when there is nearly no correlation between the indicators.	Preferable only in situations with very low sample sizes and very low R².

Given that Mode A can be used in both cases, researchers might wonder how they should evaluate whether their composite are good composites and behave as expected (i.e., reflective/formative)? We recommend that researchers expecting a reflective composite based on theory should evaluate the loadings and measures based on loadings (e.g., AVE, composite reliability etc.) of these "reflective" composites just as they did before. However, caution is necessary: a composite is not a factor and thus factor-based approaches to factor validation will never be perfectly able to assess composites. In contrast, researchers expecting a formative composite based on theory should evaluate the weights[6] and other measures typically used to validate formative composites (e.g., multicollinearity assessment). In addition, they should evaluate content validity rigorously given the importance to include all relevant formative indicators in the composite. This is not necessarily important for reflective measures as indicators should strongly covary and be interchangeable.

Limitations and Future Research

While our study makes contributions, it has its limitations and opens up avenues for future research. First, the simulation in this study limits the assessment of Mode A and Mode B to parameter recovery and in-sample and out-of-sample prediction but does not assess statistical power. Future simulations should include an assessment of the statistical power of Mode A and Mode B to estimate composite weights. While we employ an exhaustive design with many design-factor constellations to especially illustrate the effect of a predictive analysis which resulted in 420,000 PLS path model computations of a Mode A and Mode B model, the assessment of statistical significance via bootstrapping (with a usual default of 1,000 resamples) would blow up the design to an infeasible 420,000,000 model estimations of both Mode A and Mode B, and would be difficult to report within the space constraints. Second, we used a simple model with only two exogenous and one endogenous composite. Future simulations should also investigate more complex models with multiple interrelated endogenous composites. Third, we used only one set of weights

---

[6] Weights and loadings are proportional in Mode A in that the loadings represent the bivariate correlations between indicator and composite and weights are scaled correlations to produce unit variance of the composite. However, these scaled correlation weights are better comparable to regression weights derived from a Mode B estimation than the loadings and should thus be more intuitive and better comparable to older PLS studies where Mode B was usually employed for formative composites.

which was the same for all three composites and assumed the exogenous composites to be uncorrelated. Future research should assess the effect of different weight patterns (e.g., very similar/equal true weights vs. very unequal true weights) and the effect of collinearity between the exogenous composites. Fourth, researchers should develop appropriate criteria to assess the reliability of composites that are not based on factor-methods. Further untangling composite and factor-based methods in their evaluation is a valuable step towards higher quality research with composite-based methods. Fifth, concrete performance criteria like out-of-sample prediction offer a useful basis for making methodological choices. Future simulations studies involving composite methods should routinely assess predictive abilities of the methodological choices. Sixth, comparisons of PLS path modeling and CB-SEM (composite-based vs. factor-based methods) should also include situations where the model does not comply with the assumptions of a factor-model.

Appendix A

Table A.1 Path Coefficients Estimation Error, In-Sample and Out-of-Sample R² for Different Population R² Levels.

Sample Size	R^2	Path Coefficients RMSE			In- Sample R²			Out-of-Sample R²		
		Mode B	Mode A	Unit Weights	Mode B	Mode A	Unit Weights	Mode B	Mode A	Unit Weights
40	.20	.191	.158	.136	.42	.35	.18	-.06	.04	.13
	.30	.148	.132	.131	.48	.41	.25	.07	.17	.22
	.40	.126	.118	.127	.53	.48	.32	.20	.28	.31
	.50	.113	.109	.125	.59	.55	.39	.33	.39	.40
	.60	.104	.102	.122	.65	.62	.46	.45	.50	.49
	.70	.097	.097	.122	.72	.69	.53	.57	.61	.57
	.80	.094	.094	.121	.79	.77	.60	.69	.71	.66
100	.20	.084	.078	.090	.28	.26	.16	.11	.15	.15
	.30	.072	.070	.090	.37	.34	.23	.23	.25	.24
	.40	.065	.064	.093	.45	.43	.30	.34	.36	.33
	.50	.059	.059	.095	.53	.52	.38	.44	.46	.41
	.60	.053	.054	.098	.62	.61	.45	.55	.56	.50
	.70	.050	.049	.101	.70	.70	.52	.65	.67	.59
	.80	.046	.045	.103	.79	.79	.59	.76	.77	.67
500	.20	.034	.034	.056	.22	.21	.15	.19	.19	.17
	.30	.031	.031	.063	.31	.31	.22	.29	.29	.25
	.40	.028	.028	.070	.41	.41	.30	.39	.39	.34
	.50	.025	.025	.077	.51	.50	.37	.49	.49	.42
	.60	.023	.023	.083	.60	.60	.44	.59	.59	.51
	.70	.020	.020	.088	.70	.70	.52	.69	.69	.59
	.80	.018	.018	.093	.80	.80	.59	.79	.79	.68

References

Aguirre-Urreta, M. I. and Marakas, G. M. 2012. "Differential Effects of Omitting Formative Indicators: A Comparison of Techniques," in *Proceedings of the International Conference on Information Systems*, Orlando, Florida.

Armstrong, J. S. 2001a. "Combining Forecasts," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong (ed.), New York: Springer, pp. 417-439.

Armstrong, J. S. 2001b. "Standards and Practices for Forecasting," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong (ed.), New York: Springer, pp. 679-732.

Bagozzi, R. P. 2011. "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly* (35:2), pp. 261-292.

Becker, J.-M., Rai, A., Ringle, C. M., and Völckner, F. 2013. "Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats," *MIS Quarterly* (37:3), pp. 665-694.

Becker, J.-M., Klein, K., and Wetzels, M. 2012. "Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models," *Long Range Planning* (45:5/6), pp. 359-394.

Bollen, K. A. and Davis, W. R. 2009. "Causal Indicator Models: Identification, Estimation, and Testing," *Structural Equation Modeling: A Multidisciplinary Journal* (16:3), pp. 498–522.

Bollen, K. K. and Lennox, R. R. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), pp. 305-314.

Cadogan, J. W. and Lee, N. 2013 "Improper Use of Endogenous Formative Variables," *Journal of Business Research* (66:2), pp. 233–241.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Erlbaum, pp. 295-358.

Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), pp. 189-217.

Chin, W. W., Thatcher, J., and Wright, R. T. 2012. "Assessing Common Method Bias: Problems with the ULMC Technique," *MIS Quarterly* (36:3), pp. 1003-A11.

Churchill, G. W. 1979. "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* (16:1), pp. 64-73.

Diamantopoulos, A. 2011. "Incorporating Formative Measures Into Covariance-Based Structural Equation Models", *MIS Quarterly* (35:2), pp. 335-358.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., and Kaiser, S. 2012. "Guidelines for Choosing Between Multi-Item and Single-Item Scales in Construct Measurement: A Predictive Validity Perspective," *Journal of the Academy of Marketing Science* (40:3), pp. 434-449.

Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.

Dijkstra, T. K. 2010. "Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer-Verlag, pp. 23-46.

Dana, J. and Dawes, R. M. 2004. "The Superiority of Simple Alternatives to Regression for Social Science Predictions," *Journal of Educational and Behavioral Statistics* (29:3), pp. 317-331.

Einhorn, H. J. and Hogarth, R. M. 1975. "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Performance* (13:2), pp. 171-192.

Esposito Vinzi, V., Trinchera, L., and Amato, S. 2010. "PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement," in *Handbook of partial least squares: Concepts, methods and applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Springer-Verlag, Berlin, pp. 47-82.

Evermann, J. and Tate, M. 2012. "Comparing the Predictive Ability of PLS and Covariance Models," in *Proceedings of the International Conference on Information Systems*, Orlando, Florida.

Fornell, C., and Bookstein, F. L. 1982. "Two Structural Equation Models: LISREL and PLS Applied to Exit-Voice Theory," *Journal of Marketing Research* (19:4), pp. 440-452.

Friedman, M. 1953. *Essays in Positive Economics*, Chicago: University of Chicago.

Gefen, D., Rigdon, E. E., and Straub, D. 2011. "An Update and Extension to SEM Guidelines for Administrative and Social Science Research," *MIS Quarterly*, (35:2), pp. iii-A7.

Gerbing, D. W. and Anderson, J. C. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment," *Journal of Marketing Research* (25:2), pp. 186-192.

Goodhue, D., Lewis, W., and Thompson, R. 2007. "Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators," *Information Systems Research* (18:2), pp. 211-227.

Goodhue, D. L., Lewis, W., and Thompson, R. 2012. "Does PLS have Advantages for Small Sample Size or Non-Normal Data?" *MIS Quarterly* (36:3), pp. 981-A16.

Guttman, L. 1955. "The Determinacy of Factor Score Matrices with Implications for Five Other Basic Problems of Common-Factor Theory," *British Journal of Statistical Psychology* (8:2), pp. 65-81.

Hair, J. F., Ringle, C. M., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *Journal of Marketing Theory & Practice* (19:2), pp. 139-152.

Henseler, J. and Chin, W. W. 2010. "A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Path Modeling," *Structural Equation Modeling: A Multidisciplinary Journal* (17:1), pp. 82-109.

Henseler, J., Fassott, G., Dijkstra, T., and Wilson, B. 2012. "Analysing Quadratic Effects of Formative Constructs by Means of Variance-Based Structural Equation Modelling," *European Journal of Information Systems* (21:1), pp. 99-112.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. 2009. "The Use of Partial Least Squares Path Modeling in International Marketing," in: *Advances in International Marketing*, R.R. Sinkovics and P.N. Ghauri (eds.), Emerald, Bingley, pp. 277-319.

Hwang, H., Malhotra, N., Kim, Y., Tomiuk, M., and Hong, S. 2010. "A Comparative Study on Parameter Recovery of Three Approaches to Structural Equation Modeling," *Journal of Marketing Research* (47:4), pp. 699-712.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research* (30:2), pp. 199-218.

Jöreskog, K. G. 1978. "Structural Analysis of Covariance and Correlation Matrices," *Psychometrika* (43:4), pp. 443-477.

Jöreskog, K. G. 1982. "The LISREL Approach to Causal Model-Building in the Social Sciences" in *Systems Under Indirect Observation*, Part I, H. Wold and K. G. Jöreskog (eds.), Amsterdam: North-Holland, pp. 81-100.

Lee, N., Cagogan, J. W., and Chamberlain, L. 2013. "The MIMIC Model and Formative Variables: Problems and Solutions." AMS Review (3:1), pp. 3-17.

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

Lu, I. R., Kwan, E., Thomas, D., and Cedzynski, M. 2011. "Two New Methods for Estimating Structural Equation Models: An Illustration and a Comparison with Two Established Methods," *International Journal of Research In Marketing*, (28:3), pp. 258-268.

Makridakis, S., and Winkler, R.L. 1983. "Averages of Forecasts: Some Empirical Results," *Management Science*, (29:9), pp. 987-996.

Makridakis, S., and Hibon, M. 2000. "The M3 Competition: Results, Conclusions and Recommendations," *International Journal of Forecasting* (16:4), pp. 451-476.

Maraun, M. D. 1996. "Metaphor Taken as Math: Indeterminacy in the Factor Analysis Model," *Multivariate Behavioral Research* (31:4), pp. 517-538.

McDonald, R.P. 1999. *Test Theory: A Unified Treatment*. Mahwah, NJ: LEA.

Messick, S. 1989. "Validity," in *Educational Measurement (3rd ed.)*, R. L. Linn (ed.), New York: American Council on Education and Macmillan, pp. 13-103.

Mulaik, S. 2010. *Foundations of Factor Analysis (2nd ed.)*. Boca Raton, FL: Chapman & Hall / CRC.

Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.

Qureshi, I. and Compeau, D. 2009. "Assessing Between-Group Differences in Information Systems Research: A Comparison of Covariance- and Component-Based SEM," *MIS Quarterly* (33:1), pp. 197-214.

Reinartz, W. J., Haenlein, M., and Henseler, J. 2009. "An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM," *International Journal of Research in Marketing* (26:4), pp. 332-344.

Rigdon, E.E., 2012a. "Comment on "Improper use of endogenous formative variables"," *Journal of Business Research*, http://dx.doi.org/10.1016/j.jbusres.2012.08.005

Rigdon, E. E. 2012b. "Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods," *Long Range Planning* (45:5/6), pp. 341-358.

Rigdon, E. E. 2013a, "Lee, Cadogan and Chamberlain: An Excellent Point . . . But What about that Iceberg?" AMS Review, (3:1), pp. 24-29.

Rigdon, E. E. 2013b. "Partial Least Squares Path Modeling," in Structural Equation Modeling: A Second Course (2nd ed.), G. Hancock and R. Mueller (eds.), Charlotte, NC: Information Age, pp. 81-116.

Ringle, C. M., Sarstedt, M., and Straub, D. 2012. "A Critical Look at the Use of PLS-SEM in MIS Quarterly," *MIS Quarterly* (36:1), pp. iii-viii.

Rönkkö M., and Evermann J. 2013. "A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling," *Organizational Research Methods* (16:3), pp. 425-448.

Schönemann, P.H. and Haagen, K. 1987. "On the Use of Factor Scores for Prediction," *Biometrical Journal* (29:7), pp. 835-847.

Schönemann, P.H. and Steiger, J.H. 1978. "On the Validity of Indeterminate Factor Scores," *Bulletin of the Psychonomic Society* (12:4), pp. 287-290.

Shadish, W. R., Cook, T. D., and Campbell, D. T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Belmont, CA: Wadsworth Cengage Learning.

Sharma, P. N., and Kim, K. H. 2012. "Model Selection in Information Systems Research Using Partial Least Squares Based Structural Equation Modeling," in *Proceedings of the International Conference on Information Systems*, Orlando, Florida.

Shmueli, G. and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly*, (35:3), pp. 553-572.

Steiger, J. 1979a. "Factor Indeterminacy in the 1930's and the 1970's Some Interesting Parallels," *Psychometrika* (44:1), pp. 157-167.

Steiger, J. 1979b. "The Relationship between External Variables and Common Factors," *Psychometrika* (44:1), pp. 93-97.

Steiger, J. H. and Schönemann, P.H. 1978. "A History of Factor Indeterminacy," in *Theory Construction and Data Analysis in the Social Sciences*, S. Shye (ed.), San Francisco: Jossey-Bass, pp. 136-178.

Velicer, W. F. and Jackson, D. N. 1990. "Component Analysis Versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure," *Multivariate Behavioral Research* (25:1), pp. 1-28.

Wainer, H. 1976. "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin* (83:2), pp. 213–217.

Waller, N. 2008. "Fungible Weights in Multiple Regression," *Psychometrika* (73:4), pp. 691-703.

Waller, N. and Jones, J. 2010. "Correlation Weights in Multiple Regression," *Psychometrika* (75:1), pp. 58–69.

Wold, H. 1982. "Soft Modeling: The Basic Design and Some Extensions," in *Systems Under Indirect Observations: Part I*, K. G. Jöreskog and H. Wold (eds.), Amsterdam: North-Holland, pp. 1-54.

Wold, H. 1985. "Partial Least Squares," in *Encyclopedia of Statistical Sciences, Vol. 6,* S. Kotz and N. L. Johnson (eds.), New York: Wiley, pp. 581-591.

Wold, H. 1988. "Specification, Predictor," in *Encyclopedia of Statistical Sciences, Vol. 8,* S. Kotz and N. L. Johnson (eds.), New York: Wiley, pp. 587-599.

Yong, E. 2012. "Replication Studies: Bad Copy," *Nature* (485: 7398), pp. 298-300.