Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

CONSTRUCT VALIDITY IN PARTIAL LEAST SQUARES PATH MODELING

Mikko Rönkkö
Aalto University, mikko.ronkko@tkk.fi

Jukka Ylitalo
Aalto University, jukka.ylitalo@tkk.fi

CONSTRUCT VALIDITY IN PARTIAL LEAST SQUARES PATH MODELING

Completed Research Paper

Mikko Rönkkö
Aalto University, School of Science and Technology
PO Box 15500, FI-00076 Aalto
mikko.ronkko@tkk.fi

Jukka Ylitalo
Aalto University, School of Science and Technology
PO Box 15500, FI-00076 Aalto
jukka.ylitalo@tkk.fi

Abstract

Partial least squares path modeling (PLS) has seen increased use in the information systems research community. One of the stated key advantages of PLS is that it weights the indicator variables based on the strength of the relationship between the indicators and the underlying constructs, which presumably decreases the effect of measurement error in the analysis results. In this paper we argue that this assumption is not valid. While PLS indeed does weight the indicators to maximize the explained variance, it does this by including error variance in the model thus reducing construct validity. We use a simulation study of a simple PLS model to show that when compared to traditional sum scale approach, PLS estimates are actually often less valid. Although our study has its limitations, it hints that the use of PLS as a theory testing tool should be re-evaluated and that more research testing the effectiveness of the PLS approach is in order.

Keywords: Partial least squares, construct validity, Monte Carlo simulation

Introduction

Partial least squares path modeling (PLS) has seen increased use among information systems researchers (Marcoulides and Saunders, 2006). Most commonly, this method is used as an alternative to structural equation modeling (SEM) when evaluating path models; These models consists of several constructs that are hypothesized to be causally linked to form a structural model and each construct is measured with one or more indicator variables. PLS differs fundamentally from SEM by dealing with composites[1] rather than treating the hypothetized constructs as latent variables (Marcoulides, Chin, and Saunders, 2009; McDonald, 1996). In the PLS approach, the composites are considered as approximations of latent constructs, and the structural model consists of regression paths between the composites. Although in the IS literature, researchers often choose between PLS and SEM as their main analysis method, this is a false dichotomy: The fundamental question in analyzing these models is whether one wants to treat the constructs as latent variables or estimate scores for the constructs. If one chooses to not calculate scores for the constructs, SEM with latent variables is the most appropriate choice. However, if the one decides to calculate scores for the constructs and then estimate the hypothetized relationships between the constructs with regression analysis, PLS is one but not the only available alternative (McDonald, 1996).

In the PLS approach, the composites are considered as approximations of the constructs, and the structural model consists of regression paths between the composites. The justification for using PLS rather than "first generation" alternatives (Fornell, 1985) relates to measurement. Most commonly this is presenting by discussing the ability of SEM to explicitly model and partial out measurement error and implicitly stating that PLS would do the same by presenting PLS as a method for estimating SEM models (Chin, 1998; Haenlein and Kaplan, 2004; Henseler, Ringle, and Sinkovics, 2009). Sometimes the argument is presented more explicitly by stating that the weighting of the items in PLS minimizes the effect of measurement error (Chin, Marcolin, and Newsted, 2003). Earlier papers about PLS argued that the fact that the item blocks are analyzed in the context of the theoretical model would provide an advantage (Fornell, 1985; Thompson and Higgins, 1991). And sometimes it is just stated that the PLS results are not affected as strongly by measurement error as regression estimates (Pavlou, 2002) or that PLS separates out irrelevant variance from the structural model (Fornell and Bookstein, 1982) without explaining how the method accomplishes this.

The most straightforward way estimate path models of constructs can be called traditional path analysis with simple summed scales (Chin, 2010) In this approach composite variables are created by calculating a sum of the indicators[2] and then running separate regression for each endogenous variable[3] using the calculated scores. This way each indicator variable has equal contribution to the values of the composite. This approach has the advantage that all the statistical theory without latent variables applies immediately without any modifications (McDonald, 1996). Thus after forming the composites, parameters for the path model may be estimated the same way as dealing with regression equations for endogenous variables. We will compare the similarities and differences between this approach with PLS later in the paper.

Our main argument in the paper is that the indicator weightings by PLS, while indeed maximizing the explained variance in the data, results in biased estimates and do not really control for measurement error making this potentially worse alternative than the traditional path analysis approach for estimating path models of constructs. The reason for this feature is that instead of attempting to minimize the effect of error variance in the model estimates, the PLS algorithm actually tries to explain this variance with the model. The main contribution of our paper is a simulation study that clearly indicates that in most cases partial least squares path modeling results in lower construct validity than estimating the same model and data with summed scales and regression analysis. The results hint that the use of the PLS path modeling as a tool for statistical hypothesis testing should perhaps be reconsidered, although more analysis on the PLS method is clearly in order. While PLS can handle both reflective

---

[1] "A composite random variable is a function – usually a weighted sum – of component random variables, and is then observable if and only if each of its components is observable" (McDonald, 1996: 239).

[2] The indicators are often standardized before summing to eliminate effects of different measurement scales and the sum is often standardized as well to make it easier to interpret the results.

[3] Endogenous variables are determined theoretically by factors within the model, while exogenous variables are determined by factors outside the model. Thus exogenous variables are not explained by any other variable in the model while endogenous variables have at least one explanatory variable.

and formative measurement, we focus on reflective measures since these are more commonly used in IS (Cenfetelli and Bassellier, 2009). Moreover, contemporary psychometrics considers formative measurement problematic (Howell, Breivik, and Wilcox, 2007) and the traditional measures of validity do not apply to formative measurement (Bollen, 1989: 222-223) and hence these would require an entirely different set of analyses.

The rest of the paper is structured as follows: We start by presenting an overview of construct measurement and discuss the concepts of validity and reliability. Then we move to discussion on how path models can be analyzed and particularly PLS of these methods deals with measurement error. After this review of the existing theories and methods, we present our simulation design for testing how well PLS can recover true construct values of a very simple structural model. Thereafter, in the results section we show that the use of PLS results in less accurate estimates of construct values than what would be obtained with summed scales. Finally, we present conclusions including limitations and suggestion for further analysis.

Overview of Construct Measurement

Since the focus of the paper is on validity of measurement, we start by first discussing statistical hypothesis testing in general and focusing on reliability and validity in particular. After this discussion, we build our case that the PLS method, although valuable for prediction, can potentially provide misleading results when used for theory testing; Not all analysis methods are equally good for different task and what is considered as valid statistical analysis depends on the goal. Hence a refresher on statistical prediction and statistical explanations is in order.

The tasks of statistical analyses can be divided into two categories: statistical prediction and statistical explanation. Prediction is concerned with finding a model that best predicts the criterion variable regardless of the cause. We should bear in mind that this is what PLS was developed for (Wold, 1982) and this is where it is suggested to shine (Chin, 1998). The main yardstick on which the quality of a prediction model is judged is criterion related validity, or how well the predicted values correlate with the observed ones regardless of the underlying causal mechanisms (Nunnally, 1978). For explanatory models, judging the quality of the model is not as straightforward, since one needs to first consider the reliability and validity of the measured variables, then test the strength of the relationships between these variables while also controlling for any alternative explanation, and finally evaluating how well the model tested with one set of data applies to other sets. Since statistical analysis in information system is mostly used as a tool for theory testing instead of predicting the values of some criterion variable for the cases, the criteria for reliability and validity should be chosen according to the statistical explanation approach.

Most of the time the concepts that we build our theories from are abstract. Since abstract concepts are not directly measurable, they are often measured by several items that form a construct (Borsboom, Mellenbergh, and van Heerden, 2003; Nunnally, 1978, p. 96). The degree to which our empirical measures reflect the latent constructs they are supposed to measure is termed construct validity. A key problem with estimating the goodness of the measurement is that we cannot know the true scores and hence cannot directly measure how well our measures correlate with the true score. The traditional true score measurement theory states that our measured scores are a result of the true score and measurement error. In the most traditional form, this is expressed as

$$x = t + e \tag{1}$$

where x is the measured score, t is the true score, and e is error term. In this basic form the error term is considered to be a result of error in reliability. That is, a result of random variations in measurement instead of any systematic error. Since error is considered to be random, we can estimate reliability by using multiple items as parallel tests and then based on the correlations between the items estimate the mean correlation between the items and the true score (Nunnally, 1978, p. 203). The problem with this approach is that although it provides a measure of reliability, it assumes that all reliable variance is related to the construct, while in reality the responses on the items can be caused by several different factors of which only one is the construct. Hence this equation is often expressed in the following form:

$$x = t + s + e \tag{2}$$

where x is the measured score, t is the true score, s is item specific and reliable error component, and e is random error. Assuming that measured trait or aspect is relatively static, repeating the measurement would result in identical t and s but different e for each case. Hence we consider t and s to be reliable although only the first one of these is "correct" variance. In other words, since s contains all the variance not related to the construct, this, although reliable in the meaning that it is not just random noise, is still considered as error variance from the perspective of construct measurement and hence we call this component reliable error variance. If s correlate across items we would consider this systematic measurement error that can manifest e.g. as common method bias. However, often s contains mainly the unique nature of the item.

Consider the following item where an informant is asked to rate his agreement on a scale from one to seven: “Our personnel are committed to implementation of new information system”. Let us for the sake of giving an example consider that the value that best reflects the true commitment of the personnel of a focal organization would be '5'. A person responding to the question can respond to anything between three to seven depending on her personal factors or mood during responding to the questionnaire (P. M. Podsakoff, MacKenzie, Jeong-Yeon Lee, and N. P. Podsakoff, 2003). If this item was used to measure a construct with a number of similar items that were also susceptible to the same person effect, then from the perspective of parallel test reliability, the correlated error terms would manifest as reliable variance from the traditional test theory perspective and would thus influence the estimate of the true score for the particular construct. Since correlated error terms do cause bias in the construct values, psychometricians have developed techniques to control measurement error. Generally the key idea in these approaches is to partial out components not related to the measured construct from the indicators (Cronbach, Gleser, Nanda, and Rajaratnam, 1972; see Rentz, 1987).

Measurement error can cause two types of problems: Random error in the constructs decreases the statistical power of the test increasing Type II statistical error. In other words, if not controlled for, noise in the data can cause a researcher to conclude that a relationship does not exists in the population while in reality it does. Since a result of not finding an effect in statistical analysis is often not considered as sufficient evidence of non-existence of the effect in real life, but attributed to measurement error, this type of errors are often not serious. Type I error of falsely concluding that a relationship exists while it in fact does not, is a much more serious result, since this can lead to publishing results that are not valid. The problem is that both Type I error and Type II error can results when the error terms of indicators of two different constructs correlate, as illustrated by a recent article about method variance (Lance, Dawson, Birkelbach, and Hoffman, 2010).

To control for the effects of less than perfect reliability of indicators, at least three different approaches exist for estimating the relationships between the constructs: (1) using composite variables with prior weights, (2) using composite variables with data-dependent weights, or (3) using latent variables (McDonald, 1996). The idea of the first two approaches relies on the assumption of measurement error being random or close to random and thus combining several items with error cancels out part of this noise in the data. In contrast to the first two manifests variables approaches, the third, latent variable approach is a class of methods where the construct values are not estimated at all as a part of the model estimation, but the model parameters are estimated indirectly from the covariances between the indicators. Structural equation modeling with maximum likelihood estimation is perhaps the most widely known and applied of these methods.

Overview of PLS and Comparison with Alternatives

Partial least squares estimation does construct measurement by approximating the true values of the constructs with weighted sums of the indicators and then using the values of these composites to estimate the path coefficients of the model. We prefer to think this as a three step process: In the first step we estimate the indicator weights and composite scores, in the second step we run a separate OLS regression analysis on each endogenous variable to estimate the path coefficients between the constructs, and in the third step these two steps are replicated with several bootstrap samples to estimate standard errors for the model parameters. Calculating the indicator weights is a core feature of PLS and hence deserves more attention.

The iterative estimation of weights proceeds by repeating so-called inner and outer estimation steps until the model parameters have converged to the final values. The process starts by replacing each construct value with the unweighted average of its standardized indicators. Then in the inner estimation stage these construct values are used to estimate the coefficient between the constructs. This involves running a separate OLS regression on each endogenous construct. The next step is to estimate new values for the constructs. A couple of different weighting schemes exist, but generally this involves calculating the values of each construct as a linear combination of all constructs that it is linked to with a path and weighting these based on the results of the regression analyses. Then we proceed to outer estimation where we regress each indicator on the calculated construct values. The values of these regression paths are used to determine new indicator weights and a new set of construct scores are calculated as a weighted average of the indicators. The inner and outer estimation are repeated until the changes to the construct scores become sufficiently small so that we can consider the model as converged. After this, we use construct values after the final outer estimation as the final estimates for construct scores.

When compared this with the simple summed scales approach, estimating a model with PLS differs in three aspects: First, instead of treating each indicator as an equally good indicator by imposing equal weights, PLS weights the indicators to iteratively maximize the variance explained by the measurement model and structural model. Second, the standard errors for path coefficients are estimated with bootstrapping the entire model instead of relying on the standard errors of each regression separately. Third, current software implementations of PLS hide some of the details of estimating and interpreting the regression models. Although the latter point is not strictly speaking a feature of the PLS algorithm, we believe that this 'convenience factor' affects the choice and use of the method. Table 1 summarizes three methods for estimating a path model. (See Gefen, Straub, and Boudreau, 2000 for a similar comparison.)

Table 1: Comparison of different estimation approaches for path models

	PLS	**Summed scales and regression**	**SEM**
Estimating construct values	Construct values are estimated as weighted linear combinations of indicators.	Construct values are estimated as unweighted linear combination of indicators.	Construct values are not estimated. Constructs are treated as latent variables.
Calculating path coefficients between the constructs	Separate OLS regression is run for each endogenous variable.	Separate OLS regression is run for each endogenous variable.	All model coefficients are typically estimated by constructing a model implied covariance matrix and then iteratively minimizing a discrepancy function between the this and observed covariance matrix.
Estimating standard errors for the model parameters	Standard errors are estimated by resampling the entire model. (Standard errors based on known probability distribution are possible[4])	Standard errors are typically estimated based on known probability distributions. (Resampling estimation is possible.)	Standard errors are estimated based on known probability distribution.

A popular argument that is used to justify the use of PLS method instead of alternatives is that the indicator weighting scheme gives more emphasis to indicators that are the most reliable measures of the latent constructs (Chin et al., 2003). This relates to the fact that PLS has an implicit assumption that all measured variance of the variables in the model is useful variance that should be explained, but in reality this assumption often does not hold as explained in the example above. In fact, the PLS estimates of construct values are unbiased only under the unrealistic condition of non-correlated error terms (Dijkstra, 1983; McDonald, 1996). For this assumption to hold, s and e in equation 1 should be uncorrelated across all indicators in the sample. Since we know that even if the error terms would not be correlated in the population, the correlations of these random variables follow a known sampling distribution, it is very unlikely that they all would be exactly zero in the sample. However, we are not aware of any studies that test the effect of correlated error terms for PLS estimates and although the presence of correlated errors is relatively easy to test by for example examining the magnitude of crossloadings in item to construct crossloading matrix, we have not seen this applied in any PLS paper. For SEM the standardized root mean square residual fit index described the magnitude of residual covariances after the model implied covariances are partialled out from the data. One rule of thumb indicates that .10 would be the highest acceptable level of miss fit (Kline, 2005, p. 141), which implies that if all the error correlations were on equal level, the highest acceptable level would be .316. Based on this, we can conclude that the error correlations probably typically vary between .3 and -.3 for well fitting SEM models.

[4] e.g. Qureshi & Compeau, 2009

While the weighting scheme iteratively minimizes the sum of squares of residuals in each regression analysis, it does not immediately follow that the measurements would be less erroneous than if the items were weighted equally. If we were to estimate the strength of a relationship between two constructs, an ideal composite variable method would first establish a set of scores that as closely as possible reflects the true scores and then use these measured scores to estimate the relationship between the constructs. In contrast to this ideal approach, PLS method uses the information from the inner model to estimate the construct values thus departing from this ideal approach. In other words, PLS forms empirical constructs that maximize the explained variance in the inner model (thus maximizing the path coefficients) but this does not necessarily mean that the correlation between the true values of the constructs and the composites would be maximized. Even if PLS estimates are consistent under the joint condition of large sample size and large number of indicators, "consistency does not imply that the estimator is unbiased in the limit" (Wold, 1980, p. 67)

We can illustrate biased estimates with an example. Consider a model with two constructs A and B having three indicators each having equal loadings on the construct. Consider also that the first indicators of both constructs have error components that are positively correlated while the rest of the error components are uncorrelated in the sample. There is a regression path from A to B. The population model is shown in Figure 1 and we assume that the sample variance-covariance matrix is equal to the population variance-covariance matrix.

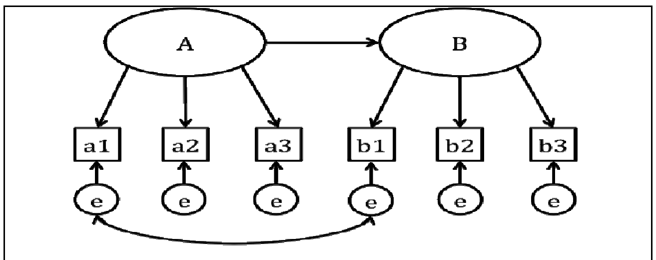

Figure 1: Example model with correlated error term

If analyzed with standardized summed scales, we just combine all variance in the constructs and estimate the regression between these composites. With this method the estimate of the correlation between A and B is biased, since each construct is estimated as a sum of the true score components A and B and all measurement error components. Since the standardized regression coefficient for a two-variable regression is the correlation between these variables, we can present the standardized regression coefficient as

$$\beta = \frac{Cov\left(\sum_{i=1}^{3} a_i, \sum_{i=1}^{3} b_i\right)}{\sqrt{Var\left(\sum_{i=1}^{3} a_i\right) Var\left(\sum_{i=1}^{3} b_i\right)}} \tag{3}$$

After we apply the calculation rules for covariances, we get

$$\begin{gathered} \beta = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3} Cov(a_i, b_j)}{3+6\lambda^2}, \text{ where} \\ Cov(a_i, b_j) = \lambda^2 Cov(t_a, t_b) \text{ if } i \neq 1 \wedge j \neq 1 \text{, and} \\ Cov(a_1, b_1) = \lambda^2 Cov(t_a, t_b) + Cov(e_{a1}, e_{b1}) \end{gathered} \tag{4}$$

Which can be simplified as

$$\beta = \frac{3\lambda^2}{1+2\lambda^2} Cov(t_a, t_b) + \frac{1}{3+6\lambda^2} Cov(e_{a1}, e_{b1}) \tag{5}$$

In the equation 5 we can see how the measurement error attenuates the regression coefficient making this an inefficient estimator for the path coefficient in the population and how the error correlation causes bias in the positive direction. The coefficient that is used to multiply the true covariance is the attenuation effect and the second term in the sum is the positive bias caused by the unmodeled error correlation.

The formula for the regression coefficient between A and B as estimated by PLS can be derived from the equation 3 by adding weights. This can be simplified to

$$\beta = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3} w_{ai} w_{bj} \lambda^2}{\sqrt{(\sum_{i=1}^{3} w_{ai}^2 + 2\lambda^2 (w_{a1} w_{a2} + w_{a1} w_{a3} + w_{a2} w_{a3}))(\sum_{i=1}^{3} w_{bi}^2 + 2\lambda^2 (w_{b1} w_{b2} + w_{b1} w_{b3} + w_{b2} w_{b3}))}} \times Cov(t_a, t_b) + \frac{w_{a1} w_{b1}}{\sqrt{(\sum_{i=1}^{3} w_{ai}^2 + 2\lambda^2 (w_{a1} w_{a2} + w_{a1} w_{a3} + w_{a2} w_{a3}))(\sum_{i=1}^{3} w_{bi}^2 + 2\lambda^2 (w_{b1} w_{b2} + w_{b1} w_{b3} + w_{b2} w_{b3}))}} \times Cov(e_{a1}, e_{b1}) \quad (6)$$

Again, we can see that the estimate for beta is inefficient (attenuated) and positively biased. Although it might not be immediately apparent, the coefficient of attenuation is minimized when all weights are equal. The proof is omitted due to the space constraint, but this feature becomes clear after considering the fact that in this case all indicators were equally good measures of the construct and hence there is no reason to weight one more than others. Moreover, setting a weight of an indicator to zero or close to zero would effectively mean reducing the number of indicators. However, rather than focusing on the coefficient of attenuation, we are more interested in what happens to the bias when weights are added for the indicators. We can write the coefficient multiplying the error covariance in a different form by canceling the weight of a1 from the subexpression on the third line of equation 6.

$$\frac{w_{b1}}{\sqrt{(1 + \frac{(w_{a2}^2 + w_{a3}^2 + 2\lambda^2 (w_{a1} w_{a2} + w_{a1} w_{a3} + w_{a2} w_{a3}))}{w_{a1}^2}) B}}, \text{ where} \quad B = \sum_{i=1}^{3} w_{bi}^2 + 2\lambda^2 (w_{b1} w_{b2} + w_{b1} w_{b3} + w_{b2} w_{b3}) \quad (7)$$

The numerator in the expression 7 is now independent of the weight of a1 and the denominator decreases when weight of a1 increases. This shows that if the weights of the variables with the error correlations are increased, the bias increases. Alternatively to this formal presentation, the same conclusion can be drawn intuitively by considering what happens when the weight of one or both of the variables with the error correlation is set to zero: this variable and the error correlation would be removed from the model.

A key problem with PLS in this particular example is that it the algorithm actually weights the indicators a1 and b1 more. The reason for this is that during the first inner estimation of the model the initial value of the A construct is used as a proxy for the B construct and the other way around. The covariance between one indicator of A between the current proxy estimate of B is shown in equation 8. The second and third indicators are correlated with the proxy estimates since they are caused by the correlating constructs, but the first indicators are correlated more since they have additional correlation due to correlated error variances. When the algorithm proceeds, this bias is amplified. We have switched from covariances to correlations to avoid the need for explicit consideration of standardization of the proxy values.

$$Cor(a_i, B) = \sum_{j=1}^{3} w_{bj} Cor(a_i, b_j) \quad (8)$$

The expressions for coefficients for the weights of A indicators on round n and the weights of B indicators on round n-1 are shown in equation set 9 and can be combined to form equation 10. From this equation we can show using analytical induction that if the starting weights are equal, the weights of indicators having larger covariances will converge to larger values than weights of indicators that correlate less with the other set of indicators. The proof for this is omitted due to the space constraints. Since the average correlation of first indicator of the A construct with the indicators of the B construct is higher than the rest of the A indicators due to the presence of the error correlation, the first indicators of both constructs are indeed weighted more hence amplifying the bias created by the error correlation.

$$w_{ai\ n+1} = Cor(a_i, B_n) = \sum_{j=1}^{3} w_{bj\ n}\ Cor(a_i, b_j)$$
$$w_{bj\ n} = Cor(b_i, A_{n-1}) = \sum_{i=1}^{3} w_{ai\ n-1}\ Cor(a_i, b_j) \tag{9}$$

$$w_{ai\ n+1} = Cor(a_i, B_n) = \sum_{j=1}^{3} \sum_{i=1}^{3} w_{ai\ \ n-1} Cor(a_i, b_j)^2 \tag{10}$$

We omit the structural equation modeling approach for solving the regression coefficient for two reasons. First, the model is over-identified and hence there is no analytical solution for the 21 structural equations that define the model implied covariance matrix for this model. Solving this system would require using any of the available numerical estimators for SEM. Second, the focus of this paper is on comparing summed scales and PLS, not PLS and SEM.

From this simple example, we can generalize that in the presence of correlated error terms, which is practically always the case at least to some extent, summed scales approach ignores the error while PLS amplifies the error. Since the unique error variance attenuates the relationship between the constructs but amplified correlated error variance increase the perceived relationship between the constructs, the net result is that a researcher can fool herself into thinking that measurement error is controlled, while in reality it is not. In the next section of the paper we will use Monte Carlo simulation to test how this bias of amplifying error correlations affects the results of PLS analysis under different sample size and measurement quality.

Simulation Study

Methods

Overall, our analysis approach follows a four-step approach. We start by generating a large set of data based on a population model and varying how good indicators our data are of the constructs. In the second stage, we use PLS and summed scales to approximate the values of the constructs by using the composites formed of the indicators. In the third stage, we tested how strongly the composite values calculated with these two approaches were related to the true values (reliability in classical test theory meaning) and if they were related to other constructs after the correct construct was controlled for (discriminant validity). The ultimate purpose is to test which of the two approaches (PLS and summed scale) provides most efficient estimates (that correlate more strongly with the true score) and least biased estimates (that are least affected by other constructs scores). We will explain this setting in more detail below.

To test if the PLS weights provide more valid measurements than summed scales, we used a variation of Monte Carlo simulation approach. In practice, this means using simulated data to examine the properties of the distributions of random variables (Paxton, Curran, Bollen, Kirby, and Chen, 2001) allowing researchers to assess the finite sampling performance of statistics by creating controlled conditions from which sampling distributions of parameter estimates are produced. Knowledge of the sampling distribution is the key to the evaluation of the behavior of a statistic. When performing Monte Carlo simulation in path modeling context, the researcher first creates a model with known population parameters (i.e., the values are set by the researcher). Several repeated samples are drawn from that population, and the parameters of interest are estimated for each sample. After that, a sampling distribution is estimated for each population parameter by collecting the parameter estimates from all of the samples. The properties of the sampling distribution, such as its mean or variance, are obtained from this estimated sampling distribution. Thus Monte Carlo simulation can be considered as a “brute force” approach to empirically evaluating statistics. The need for the large number of samples is due to the nature of statistical sampling and frequentist statistics: Typically the significance of an effect is estimated by the likelihood of drawing a sample where an effect is present when a similar effect does not exist in the population. Since the measured effects vary across each sample of the same population, estimating the performance of a statistical test requires drawing multiple samples and evaluating the distribution of the estimated parameters. In a typical Monte Carlo analysis, the researcher replicates the analysis several times under different conditions to identify when a statistic performs well. Similar Monte Carlo studies are common in IS for testing the performance of PLS (Qureshi and Compeau, 2009).

Latent variables are by definition unobservable as they cannot have any quantitative, empirically measurable values (Borsboom et al., 2003). Therefore we utilized an approach in which the true values of the latent variables are *a* *priori* known: First, we drew two random variables from multivariate normal distribution with unit variance. Then, we calculated a third variable that was 0.3 times the sum of these two variables and added a random error term such that the variance of the third variable was one. Since all latent variables have equal variance and the exogenous variables are uncorrelated, this results in R^2 of little less than 0.20, which is typical for models presented in IS journals. The population variance of each of these variables – that we label constructs – is one and the population mean zero. This setting is equal to a path model with two exogenous variables and one endogenous variable with regression paths of strength 0.3 between the constructs.

After establishing the construct level (structural) model, we generated values for several reflective indicator variables for each of the construct variables. This was done by multiplying construct values with factor loadings and then adding error correlations and error variance to make the sample variance of indicators equal to one[5].

The next step was to determine the conditions to vary in data generation. In total, we used five varying conditions: sample size, number of indicators, average sizes of factor loadings, variances of factor loadings, and levels of error correlations. Sample size was parametrized as having three different values: 50, 100, and 250. The minimum was set to as low as 50 because PLS is often argued not to impose restrictions on the sample size, which is in turn often used as a rationale for using PLS in situations with low sample size (Marcoulides and Saunders, 2006). Numbers of indicators were set to 3, 4, and 6 to reflect a typical PLS paper. Varying this is important as PLS is known to confine to what is called as *consistency at large*, meaning that the estimates are consistent when both sample size and number of indicators become infinite (Jöreskog and Wold, 1982).

Average sizes of factor loadings were set to three levels: 0.4, 0.6, and 0.8. In addition to average factor loading size, in order to simulate the fact that indicators do not always load equally to constructs, we parametrized the loadings by adding random variation in each of the loadings. Addition of variation in the loadings is important due to the fact that in case of equal loadings, the alleged advantage of PLS – that is, to weight the more reliable indicators more than others – becomes useless as in this case using the sum scales with equal weights is the best option since no indicator is better than another. To accommodate this into our analysis, instead of using the three specified factor loading levels, we drew the factor loadings from a uniform distribution whose mean was the factor loading level and absolute maximum deviation had three levels of 0.0, 0.1, and 0.2, which we labeled as small, medium, and large respectively. For example in a case of average factor loading of 0.6 and variation of 0.1, the factor loadings are drawn from the uniform distribution between 0.5 and 0.7. Furthermore, the error terms were further assigned to have a specific level of error correlation with the other error terms in the model as this is often the case in practical situations. We used three levels of error correlations: 0.0, 0.2, and 0.4 and drew the correlations randomly from uniform distribution where the level was the maximum absolute value of correlation. The expected value of squared correlation on the .4 level is .0533, which means that when considering the standardized root mean square residual criterion, fitting SEM model with this data would still be considered to have good fit. Hence even the most extreme of these simulation conditions is rather conservative. We replicated the analysis 500 times for each 243 unique combination of parameters, which is a fairly typical amount of replications from Monte Carlo simulation studies.

After generating the data by calculating the indicator scores, we developed two[6] operationalizations for the constructs. First, sum scale was formed as a mean of the indicator values loading on a construct. Then, we estimated the model with plspm R package version 0.1-4 using reflective indicators and the path weighting scheme[7]. The generated construct values were stored along with the parameters that were used to generate the data.

To ensure that the results obtained were not due to a bug in the plspm R-package, we analyzed a subset of the data using a PLS algorithm that we wrote in Stata for estimating this particular model[8]. The reason for not using any

[5] Sample variance was used instead of population variance since this approach was much more simple to implement.

[6] During development of this paper, we also calculated construct values based on orthogonal and oblique factor scores and with formative PLS. These results are not reported in detail in the paper since our purpose is to compare the summed scales approach with the reflective PLS approach. However, the factor scores that assume cross-loadings did not perform well and the formative PLS provided the least valid results since the nature of measurement was not the same as what was in the population model.

[7] We ran the analyses also with centroid and factor weighting schemes. The results of these analyses were similar to the patch weighting scheme and are thus not reported.

[8] This analysis is available from the authors upon request.

readily available PLS package is that at the time of writing none of the available packages could run analyses in batches, making estimating a large number of models impractical. In most of the cases the estimates produced in Stata agreed with PLS results. Only in rare cases where the algorithm did not converge fast, the results differed due to rounding errors accumulating in the estimating process.

After estimating the construct values using the two different methods, we compared these with the true score values with the two different methods. The analysis of the validity of the construct operationalizations started by first calculating all squared correlations between the true scores and each operationalization. The higher the squared correlation, the less error the operationalization contains. There are several reasons for using squared correlation instead of plain correlation[9]. First, in some rare cases the PLS algorithm converged to values that resulted in negative correlations between the construct values and true scores. Since this was considered to be an error in scale setting rather than an error in validity, we did not want to have the sign of the correlation affect the results. Second, the squared correlation conveniently tells how much of the construct value is explained by the true score while the rest is error. Third, using squared correlation allowed us to test for bias in the construct. Since squared correlation of two variables is equal to the R^2 when these variables are regressed on one another, we can compare how much the R^2 is increased if we add also the true scores of other constructs as predictors. If the measure for a construct has good discriminant validity, the R^2 should not increase significantly when the rest of the true scores are added to the model. The rationale for this is that in the population model, the indicators of two constructs are related only through the constructs and hence a measure formed from one set of indicators should not correlate with other constructs when the correct construct is controlled for. If there is a relationship between an operationalization of a construct and the true score of another construct, then this is an indication that the operationalization does not properly discriminate between the two constructs and hence any analysis of the relationship between the constructs is biased. This is an important test since the first measure captures only the reliability of the measurement in the classical test theory sense, but does not reveal any potential bias that could cause inflation of the model coefficients.

We analyzed these two measures of goodness of measurement – the squared correlation between the measured value and true score and change in R^2 when the two other true score values were added as predictors - first descriptively using tables and plots and finally using regressions to see which of the estimation parameters had most effect on the results. R code for generating the date, calculating the composites, and estimating the R^2 values for one replication is provided in Appendix 1.

Results

We start presenting the results by first taking a look at the ultimate measure of the quality of the measurement in the classical test theory: the squared correlation between the true score and the measured score. Table 2 presents the results of comparing squared correlation of PLS estimates with the true score values and summed scales with the true scores. The table shows each unique estimating condition and then two figures and a test of significance. The first figure shows the frequency of summed scale providing estimates that correlate more with the true score than what the PLS estimates did. The second figure is the frequency of PLS estimates correlating stronger than the summed scale estimates. These sum to 1500 since each of the 500 replications generated values for all of the tree constructs. The stars show a result of two single-tailed paired t-test of the two estimates. Stars in parentheses indicate that the alternative hypothesis that sum scales estimates correlate stronger with the true score than PLS estimates received more support than the other alternative hypothesis that PLS results correlated more with the true score values.

The results suggest that when the data are poor, PLS tends to produce less valid results than summed scales. In fact, on the first page of the table, every single condition produced more instances when summed scale performed better than PLS estimates and all t-tests are significant at $p<.001$. Only when the indicator count and sample size have the largest possible values and the factor loadings are at .6 or .8 level, PLS estimates correlate more with the true scores

[9] Typically researchers use Cronbach's alpha or composite reliability to estimate the reliability of the index. These are ways to approximate the correlation between the true score and the measured score in the absence of information about the true score values. Since in our simulation, the true scores were *a priori* known, we could use the correlation between the measure and true score directly.

Table 2 comparison of correlations of true score with PLS and sum scale estimates

	Sample size								
	50 Indicators			100 Indicators			250 Indicators		
	3	4	6	3	4	6	3	4	6
Factor loadings: .4									
Factor loading variance: 0									
Error correlations									
0	1136/364$^{(***)}$	1182/318$^{(***)}$	1215/285$^{(***)}$	1208/292$^{(***)}$	1263/237$^{(***)}$	1323/177$^{(***)}$	1285/215$^{(***)}$	1328/172$^{(***)}$	1402/98$^{(***)}$
.2	1225/275$^{(***)}$	1246/254$^{(***)}$	1259/241$^{(***)}$	1321/179$^{(***)}$	1344/156$^{(***)}$	1370/130$^{(***)}$	1400/100$^{(***)}$	1426/74$^{(***)}$	1448/52$^{(***)}$
.4	1330/170$^{(***)}$	1344/156$^{(***)}$	1351/149$^{(***)}$	1416/84$^{(***)}$	1430/70$^{(***)}$	1430/70$^{(***)}$	1451/49$^{(***)}$	1468/32$^{(***)}$	1482/18$^{(***)}$
Factor loading variance: .1									
Error correlations									
0	1052/448$^{(***)}$	1102/398$^{(***)}$	1073/427$^{(***)}$	1128/372$^{(***)}$	1106/394$^{(***)}$	1173/327$^{(***)}$	1035/465$^{(***)}$	1074/426$^{(***)}$	1136/364$^{(***)}$
.2	1145/355$^{(***)}$	1149/351$^{(***)}$	1181/319$^{(***)}$	1237/263$^{(***)}$	1228/272$^{(***)}$	1227/273$^{(***)}$	1290/210$^{(***)}$	1284/216$^{(***)}$	1272/228$^{(***)}$
.4	1314/186$^{(***)}$	1302/198$^{(***)}$	1291/209$^{(***)}$	1375/125$^{(***)}$	1390/110$^{(***)}$	1369/131$^{(***)}$	1405/95$^{(***)}$	1448/52$^{(***)}$	1434/66$^{(***)}$
Factor loading variance: .2									
Error correlations									
0	897/603$^{(***)}$	893/607$^{(***)}$	870/630$^{(***)}$	839/661$^{(***)}$	832/668$^{(***)}$	845/655$^{(***)}$	677/823$^{(***)}$	629/871$^{(***)}$	570/930$^{(**)}$
.2	1022/478$^{(***)}$	1001/499$^{(***)}$	953/547$^{(***)}$	1087/413$^{(***)}$	1012/488$^{(***)}$	964/536$^{(***)}$	1084/416$^{(***)}$	1008/492$^{(***)}$	857/643$^{(***)}$
.4	1230/270$^{(***)}$	1184/316$^{(***)}$	1113/387$^{(***)}$	1273/227$^{(***)}$	1287/213$^{(***)}$	1218/282$^{(***)}$	1325/175$^{(***)}$	1302/198$^{(***)}$	1260/240$^{(***)}$
Factor loadings: .6									
Factor loading variance: 0									
Error correlations									
0	1226/274$^{(***)}$	1274/226$^{(***)}$	1314/186$^{(***)}$	1263/237$^{(***)}$	1338/162$^{(***)}$	1376/124$^{(***)}$	1294/206$^{(***)}$	1371/129$^{(***)}$	1434/66$^{(***)}$
.2	1221/279$^{(***)}$	1281/219$^{(***)}$	1295/205$^{(***)}$	1259/241$^{(***)}$	1327/173$^{(***)}$	1372/128$^{(***)}$	1321/179$^{(***)}$	1357/143$^{(***)}$	1382/118$^{(***)}$
.4	1262/238$^{(***)}$	1304/196$^{(***)}$	1295/205$^{(***)}$	1324/176$^{(***)}$	1337/163$^{(***)}$	1358/142$^{(***)}$	1384/116$^{(***)}$	1377/123$^{(***)}$	1373/127$^{(***)}$
Factor loading variance: .1									
Error correlations									
0	1085/415$^{(***)}$	1088/412$^{(***)}$	1141/359$^{(***)}$	979/521$^{(***)}$	1078/422$^{(***)}$	1100/400$^{(***)}$	865/635$^{(***)}$	853/647$^{(***)}$	910/590$^{(***)}$
.2	1093/407$^{(***)}$	1117/383$^{(***)}$	1172/328$^{(***)}$	1085/415$^{(***)}$	1119/381$^{(***)}$	1135/365$^{(***)}$	1034/466$^{(***)}$	997/503$^{(***)}$	968/532$^{(***)}$
.4	1187/313$^{(***)}$	1176/324$^{(***)}$	1181/319$^{(***)}$	1195/305$^{(***)}$	1212/288$^{(***)}$	1202/298$^{(***)}$	1223/277$^{(***)}$	1180/320$^{(***)}$	1145/355$^{(***)}$

Continued on next page

Table 2 - continued from previous page

	Sample size								
	50 Indicators			100 Indicators			250 Indicators		
	3	4	6	3	4	6	3	4	6
Factor loading variance: .2									
Error correlations									
0	818/682$^{(***)}$	791/709$^{(***)}$	791/ 709$^{(***)}$	707/ 793$^{(***)}$	700/ 800$^{(***)}$	653/ 847$^{(***)}$	455/1045***	432/1068***	338/1162***
.2	884/616$^{(***)}$	850/650$^{(***)}$	817/ 683$^{(***)}$	803/ 697$^{(***)}$	755/ 745$^{(***)}$	715/ 785$^{(***)}$	706/ 794$^{(***)}$	549/ 951*	420/1080***
.4	1017/483$^{(***)}$	981/519$^{(***)}$	932/ 568$^{(***)}$	969/ 531$^{(***)}$	951/ 549$^{(***)}$	825/ 675$^{(***)}$	960/ 540$^{(***)}$	829/ 671$^{(***)}$	637/ 863$^{(***)}$
Factor loadings: .8									
Factor loading variance: 0									
Error correlations									
0	1246/254$^{(***)}$	1307/193$^{(***)}$	1343/ 157$^{(***)}$	1279/ 221$^{(***)}$	1344/ 156$^{(***)}$	1395/ 105$^{(***)}$	1312/ 188$^{(***)}$	1378/ 122$^{(***)}$	1434/ 66$^{(***)}$
.2	1196/304$^{(***)}$	1276/224$^{(***)}$	1307/ 193$^{(***)}$	1225/ 275$^{(***)}$	1279/ 221$^{(***)}$	1317/ 183$^{(***)}$	1215/ 285$^{(***)}$	1236/ 264$^{(***)}$	1232/ 268$^{(***)}$
.4	1166/334$^{(***)}$	1187/313$^{(***)}$	1235/ 265$^{(***)}$	1187/ 313$^{(***)}$	1174/ 326$^{(***)}$	1215/ 285$^{(***)}$	1157/ 343$^{(***)}$	1154/ 346$^{(***)}$	1123/ 377$^{(***)}$
Factor loading variance: .1									
Error correlations									
0	880/620$^{(***)}$	886/614$^{(***)}$	889/ 611$^{(***)}$	710/ 790$^{(***)}$	735/ 765$^{(***)}$	681/ 819$^{(***)}$	512/ 988***	442/1058***	406/1094***
.2	882/618$^{(***)}$	863/637$^{(***)}$	911/ 589$^{(***)}$	758/ 742$^{(***)}$	749/ 751$^{(***)}$	721/ 779$^{(***)}$	603/ 897***	514/ 986***	404/1096***
.4	883/617$^{(***)}$	888/612$^{(***)}$	915/ 585$^{(***)}$	844/ 656$^{(***)}$	797/ 703$^{(***)}$	791/ 709$^{(***)}$	697/ 803$^{(**)}$	615/ 885*	547/ 953***
Factor loading variance: .2									
Error correlations									
0	576/924$^{(***)}$	559/941$^{(***)}$	498/1002$^{(***)}$	420/1080***	378/1122*	276/1224***	233/1267***	171/1329***	87/1413***
.2	606/894$^{(***)}$	600/900$^{(***)}$	485/1015$^{(***)}$	452/1048**	389/1111**	312/1188***	291/1209***	188/1312***	99/1401***
.4	667/833$^{(***)}$	608/892$^{(***)}$	549/ 951$^{(***)}$	515/ 985†	447/1053	339/1161***	380/1120***	262/1238***	138/1362***

First number is a frequency of sum scale being more reliable and second number PLS more reliable
t-tests † $p<.1$, * $p<.05$, ** $p<.01$, *** $p<.001$ Parentheses indicate that sum scale was better.

than the sum scale estimates. However, paradoxically even if under some conditions PLS is more often better than summed scale, the t-test indicates that PLS is actually worse in that condition. The reason for this is that in the cases where PLS did provide better results, they were only marginally better and when PLS converged to poor estimates, they were substantially less correlated with the true score than summed scales.

Table 3 shows regression analyses examining what determines the relative performance of PLS algorithm over summed scales. First two equations have the difference of squared correlations as the dependent variable. If this difference is positive, then PLS provided an estimate that correlated more strongly with the true score than the summed scale estimate did. The dependent variable in the third and fourth model is the difference of change in R^2 when the rest of the constructs are added as predictors. Again, when this value is positive, then the added explanatory power of the extra constructs is smaller for PLS and hence PLS estimates have better discriminant validity than summed scales. Since the variable indicator loading has the largest coefficient of all the varied conditions, we added interaction effects of this variable with the other variables in the even models as extra predictors.

The negative intercepts suggest that on average, PLS estimates perform worse than summed scales. Also, several other features can be identified in the table. First, values of exogenous constructs are substantially less valid and more biased than endogenous construct. The reason for this is that these have only one path each and thus during the inner estimation these constructs replaced with the outer estimates of the endogenous construct thus increasing the likelihood that error correlations between these two constructs are inflated. The variance of factor loadings is positively related to the performance of the PLS algorithm suggesting that indicator weighting can indeed in some occasion weight more the items that are more reliable. All indicators of data quality are positively related to the performance of PLS model. This means that when the data is contaminated with measurement error PLS is not the optimal choice. This is in stark contrast to statements that PLS would control for measurement error.

Table 3 Determinants of bias and reliability

	PLS has better reliability		PLS is less biased	
	(1)	(2)	(3)	(4)
Indicator count	0.00294***	0.00294***	-0.00102***	-0.00102***
Factor loading mean	0.0327***	0.0213***	0.00342***	0.00265***
Factor loading variance	0.00714***	0.00714***	0.000156***	0.000156***
Sample size	0.0120***	0.0120***	0.00415***	0.00415***
Error correlations	-0.0169***	-0.0169***	0.000122***	0.000122***
Is exogenous	-0.0381***	-0.0381***	-0.00274***	-0.00274***
Indicator count x factor loading mean		-0.000294†		0.000958***
Factor loading mean x factor loading mean		-0.00535***		-0.000702***
Factor loading variance x factor loading mean		-0.00213***		-0.0000923**
Sample size x factor loading mean		0.00173***		-0.00258***
Error correlations x factor loading mean		0.0155***		-0.000184***
Is exogenous x factor loading mean		0.0170***		0.00115***
Constant	-0.0193***	-0.0139***	-0.00305***	-0.00235***
N	364500	364500	364500	364500
R^2	0.171	0.201	0.0918	0.116
F	12539.6	7636.9	6142.4	3967.6
p	0	0	0	0

† $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion and Conclusions

In the IS literature, PLS is most often compared with structural equation modeling with latent variables. However, this is a false dichotomy since PLS is but a one possible composite variables approach for path modeling. In this paper we have compared construct validity between PLS approach and the simplest composite variable approach available: the summed scales. Traditionally the use of PLS over the regular OLS regressions with summed scales has been justified by the argument that PLS comprises more valid measurements by weighting the good indicators more than others. In this paper we have demonstrated that this assumption is likely to be false in practical situations: Although PLS does weight some indicators more, it does not necessarily mean that they are the 'good' indicators that should receive more weight when maximizing construct validity. Although the rationale for data dependent weighting is probably true in ideal situations where the measures of different constructs correlate only with regard to their associated constructs and the structural paths between them, this is seldom the situation in practice. In real-life data analysis situation there are practically always error correlations in the data. In these situations, PLS – while maximizing the explained total variance of endogenous constructs – does not work as well as in ideal situations: When there is extra correlation between the items of different constructs, PLS considers these items to comprise better measurement of their associated constructs because that would maximize the capability of the structural model to explain the variance of endogenous constructs. Thus PLS weights these indicators more as they appear to be better indicators in PLS's sense, while in reality the situation is quite opposite. This weighting of the indicators falsely recognized as 'good' tends to bias the results by 'bending' the constructs toward each other.

In this paper we have provided preliminary evidence that PLS tends to bias the results when unintended inter-item correlations are present. We concentrated mostly on comparison of PLS and summed scales approach. The results suggest that in most situations, the summed scales approach does produce more valid measurement than PLS. Only under extremely favorable conditions, PLS provided better estimates than summed scales and in these cases the difference was so small as to probably not have any impact on the results. If your study generalizes from this simple model to any model, this would have weighty implications for the use of PLS: both PLS and summed scales with regression can be used to estimate path models and both have the same requirements for the data. Since the alleged advantage of PLS as producing better measurement by weighting the indicators does not seem to hold, it is unclear why this more complex method should be used instead of the simple summed scales approach.

One considerable limitation to our study is that we did not include model complexity as a parameter in our research design. Thus the results we have presented here are mostly preliminary in a sense that more research should be done on the construct validity of the PLS using different structural models. To overcome this limitation and establish the generalizability of our results further, research utilizing more versatile models and preferably real world data is required.

References

Bollen, K. A. 1989. *Structural equations with latent variables*, New York, NY: John Wiley & Son Inc.

Borsboom, D., Mellenbergh, G. J., and van Heerden, J. 2003. "The theoretical status of latent variables," *Psychological Review* (110:2), pp. 203-218.

Cenfetelli, R. T., and Bassellier, G. 2009. "Interpretation of Formative Measurement in Information Systems Research.," *MIS Quarterly* (33:4), pp. 689-707.

Chin, W. W. 1998. "Issues and Opinion on Structural Equation Modeling.," *MIS Quarterly* (22:1), pp. 1.

Chin, W. W. 2010. "Bootstrap Cross-Validation Indices for PLS Path Model Assessment," In *Handbook of Partial Least Squares*, pp. 83-97.

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: Lawrence Erlbaum Associates Publishers, pp. 295-336.

Chin, W. W. 2010. "How to Write Up and Report PLS Analyses," In *Handbook of Partial Least Squares*, pp. 655-690.

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.

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. 1972. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*, Wiley New York.

Dijkstra, T. 1983. “Some comments on maximum likelihood and partial least squares methods,” *Journal of Econometrics* (22:1-2), pp. 67–90.

Fornell, C. 1985. *A second generation of multivariate analysis: Classification of methods and implications for marketing research*, Working paper, University of Michigan.

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

Gefen, D., Straub, D., and Boudreau, M. 2000. “Structural equation modeling and regression: Guidelines for research practice,” *Structural Equation Modeling* (4:7).

Haenlein, M., and Kaplan, A. M. 2004. “A beginner's guide to partial least squares analysis,” *Understanding statistics* (3:4), pp. 283–297.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. 2009. “The use of partial least squares path modeling in international marketing,” *Advances in International Marketing* (20:2009), pp. 277–319.

Howell, R. D., Breivik, E., and Wilcox, J. B. 2007. “Reconsidering formative measurement.,” *Psychological Methods* (12:2), pp. 205-218.

Jöreskog, K. G., and Wold, H. 1982. “The ML and PLS techniques for modeling with latent variables,” In *Systems under indirect observation : causality, structure, prediction*, K. G. Jöreskog and S. Wold (eds.), Amsterdam, NY: North-Holland.

Kline, R. B. 2005. *Principles and practice of structural equation modeling*, Methodology in the social sciencesNew York: Guilford Press.

Lance, C. E., Dawson, B., Birkelbach, D., and Hoffman, B. J. 2010. “Method Effects, Measurement Error, and Substantive Conclusions,” *Organizational Research Methods*.

Marcoulides, G. A., Chin, W. W., and Saunders, C. 2009. “A critical look at partial least squares modeling,” *MIS Quarterly* (33:1), pp. 171-175.

Marcoulides, G. A., and Saunders, C. 2006. “PLS: A Silver Bullet?,” *MIS Quarterly* (30:2), pp. iii-ix.

McDonald, R. P. 1996. “Path Analysis with Composite Variables.,” *Multivariate Behavioral Research* (31:2), pp. 239-270.

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

Pavlou, P. A. 2002. “Institution-based trust in interorganizational exchange relationships: the role of online B2B marketplaces on trust formation,” *The Journal of Strategic Information Systems* (11:3-4), pp. 215-243.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., and Chen, F. 2001. “Teacher□s Corner: Monte Carlo Experiments: Design and Implementation,” *Structural Equation Modeling: A Multidisciplinary Journal* (8:2), pp. 287–312.

Podsakoff, P. M., MacKenzie, S. B., Jeong-Yeon Lee, and Podsakoff, N. P. 2003. “Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies.,” *Journal of Applied Psychology* (88:5), pp. 879-903.

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.

Rentz, J. O. 1987. “Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures.,” *Journal of Marketing Research (JMR)* (24:1), pp. 19-28.

Thompson, R. L., and Higgins, C. A. 1991. “Personal Computing: Toward a Conceptual Model of Utilization.,” *MIS Quarterly* (15:1), pp. 125-143.

Wold, H. 1982. “Soft modeling - The Basic Design And Some Extensions,” In *Systems under indirect observation : causality, structure, prediction*, K. Jöreskog and S. Wold (eds.), Amsterdam, NY: North-Holland, pp. 1-54.

Wold, H. 1980. “Model Construction and Evaluation When Theoretical Knowledge Is Scarce,” In *Model evaluation in econometrics*, J. Kmenta and J. B. Ramsey (eds.), New York, NY: Academic Press, pp. 47-74.

Appendix 1: R code for a single run of the simulation

```
require(plspm)
require(Matrix)
require(MASS)
require(ggm)

#This function runs one replication and stores it on disk. All parameters are scalars.

runtest=function(replication,sample,indicatorcount,factorloading,factorloadinginterval,errorcorrelations){

# The following matrix contains the population values for the regression coefficients of the inner model.Dependents are on rows
and independents on cols. The matrix must be square and lower triangular.
  populationmodel <- t(array(c(0,  0,  0, 0,  0,  0, .3,  .3,  0),dim=c(3,3)))

  constructcount <- nrow(populationmodel)

# GENERATE TEST DATA.

  constructs=data.frame(row.names =c(1:sample))

  for ( i in 1:constructcount ){
    newvar <- mat.or.vec(sample,1)
# Calculate fitted values if this is endogeneous variable
    for ( j in 1:constructcount){ if(populationmodel[i,j]>0) { newvar <- newvar + populationmodel[i,j]*constructs[[j]] }}

# Add residual.
    residual <- 1-var(newvar)
    newvar <- newvar + rnorm(sample)*sqrt(residual)
    constructs <- data.frame(constructs,newvar)
  }

# Calculate the indicator values
  indicators=data.frame(row.names =c(1:sample))
  residualstandarddeviations <- c()

  for ( i in 1:constructcount ){
    for ( j in 1:indicatorcount){

      thisloading <- factorloading+(runif(1)-.5)*2*factorloadinginterval
# "True values" of indicators
      indicators <- data.frame(indicators,constructs[[i]]*thisloading)

# Calculate how much residual is needed to get unit variance for indicators
      residualstandarddeviations <- append(residualstandarddeviations,sqrt(1-thisloading^2))
    }
  }

# A correlation matrix, scale it to get the maximum correlations right and then fix diagonal.
  cormat <- rcorr(indicatorcount*constructcount)*errorcorrelations+diag(1-errorcorrelations,nrow =
indicatorcount*constructcount,ncol = indicatorcount*constructcount)

# Scale the matrix so that it is a covariance matrix of the residuals that were calculated earlier
  errorcovariancematrix <- cormat * residualstandarddeviations %*% t(residualstandarddeviations)

#Draw residuals from multivariate normal distribution using the calculated matrix and add these to the indicator values. Since
error correlations are a feature of the sample, not the population, mu and Sigma are considered as sample values.
  indicators <- indicators+mvrnorm(n = sample, mu = rep(0,constructcount*indicatorcount), Sigma =
(errorcovariancematrix+t(errorcovariancematrix))/2,empirical=TRUE)

# CALCULATE CONSTRUCT VALUES WITH DIFFERENT APPROACHES

#Summed scales

  constructssum=data.frame(row.names =c(1:sample))
  for ( i in 1:constructcount ){ constructssum=data.frame(constructssum,rowMeans(indicators[((i-
1)*indicatorcount+1):(i*indicatorcount)])) }

#PLS, reflective. Start with inner and outer model matrixes.

  inner <- populationmodel!=0
  outer <- list()
  for ( i in 1:constructcount ){ outer[[i]] <- c(((i-1)*indicatorcount+1):(i*indicatorcount)) }
  constructspls <- data.frame(plspm(indicators,inner,outer, rep("A",constructcount), scheme= "path", scaled=FALSE)$scores)

# Calculate four regressions and store the r squared values
  for ( i in 1:constructcount ){
    est<-lm(as.formula(paste("constructssum[[",i,"]] ~  ",names(constructs)[[i]],sep="")) ,data=constructs)
    res1<-summary.lm(est)$r.squared
    est<-lm(as.formula(paste("constructspls[[",i,"]] ~  ",names(constructs)[[i]],sep="")) ,data=constructs)
    res2<-summary.lm(est)$r.squared
    est<-lm(as.formula(paste("constructssum[[",i,"]] ~  ",paste(names(constructs),collapse=" + "),sep="")) ,data=constructs)
    res3<-summary.lm(est)$r.squared
    est<-lm(as.formula(paste("constructspls[[",i,"]] ~  ",paste(names(constructs),collapse=" + "),sep="")) ,data=constructs)
    res4<-summary.lm(est)$r.squared

#Write the results for one construct to disk as csv

write.table(cbind(replication,i,sample,indicatorcount,factorloading,factorloadinginterval,errorcorrelations,res1,res2,res3,res4),
file="results.csv",append = TRUE, col.names = FALSE, row.names = FALSE,sep=",")
  }
}
```