Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Mathematics and Systems Analysis
Mat-2.4108 Independent Research Projects in Applied Mathematics

Controlling for Common Method Variance with Partial Least Squares Path modeling: A Monte Carlo Study

Jukka Ylitalo

64353A

11.11.2009

Contents

1. Foreword

This study is based on the paper ”Controlling for Common Method Variance with Partial Least Squares Path Modelling” that Mikko Rönkkö and I submitted for review to European Journal of Information Systems, Special Issue of Quantitative Methods on September 2009. My responsibilities with the paper involved mostly developing the presented approach, designing and implementing Monte Carlo simulation, and analysis of the obtained results. This report extends the existing analysis and includes additional information regarding, for example, structural equation modeling, partial-least squares path modeling, and Monte Carlo simulation.

2. Introduction

Use of structural equation modeling (SEM) has become pervasive among social, behavioral, and educational scientists as well as biologists, economists, marketing, and medical researchers during the recent years (Raykov & Marcoulides, 2006). The most well-known SEM techniques are methods related to covariance-based structural equation modeling (CBSEM), and some researchers use them interchangeably with the term SEM (Chin, 1998a). Besides covariance-based methods, partial least squares (PLS) path modeling has seen increased use among information systems (IS) researchers (Marcoulides, Chin, & Saunders, 2009; Marcoulides & Saunders, 2006). Due to its historical roots in principal component analysis (Wold, 1978), PLS is often also called component based approach to structural equation modeling (e.g., Qureshi & Compeau, 2009). Both of these approaches are commonly used to to estimate path models of latent variables within the IS research domain.

Analysis approaches and techniques have several fundamental differences between CBSEM and PLS. Due to these differences in model estimation, analyses approaches developed for CBSEM are not always directly applicable to PLS. Consequently, several contributions have recently been seen on how various structural equation modeling techniques can be adapted to or implemented with PLS. These include, for example, interaction effects (Chin, Marcolin, & Newsted, 2003), models with hierarchical latent constructs (Martin Wetzels, Odekerken-Schröder, & van Oppen, 2009), and multi-group models (Qureshi & Compeau, 2009).

Common method variance refers to variance that is attributable to the measurement method rather than to the constructs the measures are supposed to represent. Method biases are one of the main sources of measurement error, and most researchers agree that common method variance is a potential problem in behavioral research (P. M. Podsakoff, MacKenzie, Jeong-Yeon Lee, & N. P. Podsakoff, 2003). For CBSEM, several techniques exist for controlling for this variance, but so far few of these approaches are directly applicable to PLS. The purpose of this study is to present a PLS approach for controlling common method variance similar to single measured method factor design (P. M. Podsakoff et al., 2003) or confirmatory factor analysis (CFA) marker variable design in CBSEM (Richardson, Simmering, & Sturman, 2009). The approach is tested by using Monte Carlo simulation.

After the introduction, this study continues by introducing reader in more detail to structural equation modeling and both CBSEM and PLS approaches. This section is followed by a review of techniques for controlling common method variance. During this review, a model for controlling method variance in PLS is conceptually developed. After that, the proposed approach is tested using Monte Carlo simulation and compared to method factor design implemented with structural equation modeling. In these analyses, the technology acceptance model and the results of a recent meta-analysis by Schepers and Wetzels (2007) are used as the basis for the research model. The report is concluded by discussing several aspects, strengths, weaknesses of the proposed approach and by presenting further guidelines for diagnosing and controlling for common method variance with PLS.

3. Structural equation modeling

Structural equation modeling is a family of statistical models that seek to explain the relationships among multiple variables. In doing so, it examines the structure of interrelationships expressed in a series of equations, similar to a series of multiple regression equations (Hair, Black, Babin, Anderson, & Tatham, 2006). Structural equation modeling is an extension of the general linear model, and it has several common statistical methods as special cases, for example multiple regression, path analysis, factor analysis, time series analysis, and analysis of covariance (Maula, 2001).

The variables used in structural models are divided into observed and latent variables (Kline, 2005). Observed variables are those that can be directly measured. These can include variables like revenue, amount of personnel, and profit of the firm. Latent variables, on the other hand, cannot be directly measured: their values are estimated in the model from observed variables, called indicator variables. Latent variables are of major importance in many areas of science, and they can include factors like growth motivation of entrepreneur or goodness of firm's strategy. Both of these variables – observed and latent – are further divided into exogenous and endogenous variables. Exogenous latent variables are those variables that are not predicted by other latent variables in the model. Thus exogenous latent variables appear only as independent variables in the model equations. Endogenous latent variables, on the other hand, are predicted by some other latent variables in the model, and therefore appear as dependent variables in some of the model equations. Observed variables are divided into exogenous and endogenous variables according to which latent variables they are assigned to load on.

3.1. Covariance-based structural equation modeling

Two current main approaches to structural equation modeling are covariance-based structural equation modeling (CBSEM) and partial least squares (PLS) path modeling. Both approaches start by first specifying a path model of latent variables and then assigning a set of indicators for each latent variable. After this step, these two approaches depart. In CBSEM, the researcher traces the hypothesized factor loadings and regression paths to arrive in a set of equations describing the expected covariance structures in the data (Loehlin, 1987; Meehl & Waller, 2002). The set of equations is then used to derive a model implied covariance matrix and free parameters in the equations are estimated by minimizing the differences of the implied and observed covariance matrices.

The general CBSEM model consists of two parts: a measurement model and a structural model (Bollen, 1989). The measurement model specifies the relations of observed to latent variables. It is used to evaluate the appropriateness of the chosen indicators for estimating the latent variables, thus assessing the validity of the latent constructs. In case of a model with p endogenous and q exogenous observed variables and n endogenous and m exogenous latent variables, the measurement part of the general structural equation model is of the form

$$\begin{aligned} y &= \Lambda_y \eta + \epsilon \\ x &= \Lambda_x \xi + \delta \end{aligned} \quad , \qquad (1)$$

where $\mathbf{y}$ is $p \times 1$ and $\mathbf{x}$ is $q \times 1$ vector of observed variables, $\boldsymbol{\eta}$ is $n \times 1$ vector of latent endogenous variables, $\boldsymbol{\xi}$ is $m \times 1$ vector of latent exogenous variables, and $\boldsymbol{\epsilon}$ is $p \times 1$ and $\boldsymbol{\delta}$ is $q \times 1$ vector containing the errors of measurement (Bollen, 1989). The $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are $p \times n$ and $q \times m$ matrices containing the coefficients linking the latent and observed variables. The errors of measurement are assumed to be uncorrelated with $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ and with each other, and $E(\boldsymbol{\epsilon}) = E(\boldsymbol{\delta}) = 0$. For simplicity, $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\mathbf{y}$, and $\mathbf{x}$ are generally considered as centered, that is, written as deviations from their means. The above equations actually equivalent to confirmatory factor analyses of observed variables to their latent variables. Therefore $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ can be interpreted as matrices containing factor loadings.

The structural model encompasses the structural equations that summarize the relationships between latent variables. It shows the influence of latent variables on each other and defines exogenous and endogenous variables in the model:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad . \tag{2}$$

Here **B** is $n \times n$ matrix containing the coefficients of latent endogenous variables on each other, **Γ** is $n \times m$ matrix of coefficients between endogenous and exogenous latent variables, and ζ is the disturbance vector. It is assumed that E(ζ) = 0 and ζ is uncorrelated with ξ.

The basic hypothesis of the general CBSEM model is

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \quad , \tag{3}$$

where **Σ** is population covariance matrix of **y** and **x** and **Σ(θ)** is the covariance matrix written as a function of the free model parameters in **θ**. This implies that each element of the covariance matrix is a function of one or more model parameters. The estimation of the model thus concerns choosing the unknown parameters in **θ** so that **Σ(θ)** matches the covariance matrix of the observed variables as well as possible. Model equations in (1) and (2) are necessary to construct **Σ(θ)** according to the hypothesized model (for more details, see Bollen, 1989).

3.2. Partial least-squares path modeling

Partial least-squares (PLS) path modeling is a family of alternating least squares algorithms, or "prescriptions," which extend principal component and canonical correlation analysis (Henseler, Ringle, & Sinkovics, 2008). While in CBSEM the focus was on constructing a model implied covariance matrix and choosing the parameter estimates that minimize the difference between this and the observed covariance matrix, in PLS the purpose is to apply an iterative algorithm to directly estimate values for the latent variables.

In PLS, there are formally two model parts: the inner and outer model. These correspond to the structural and measurement models in CBSEM: the inner model specifies the relationships between latent variables, whereas the outer model specifies the relationships between latent variables and observed variables. There are two different kinds of outer models in PLS: formative and reflective. In the reflective mode, each observed variable in a certain measurement model is assumed to be generated as a linear function of its latent variables and the residual, while the formative mode has causal relationships from the observed variables to the latent variable. Due to reasons explained later, only reflective models are considered in this study.

PLS model equations can be written similarly to CBSEM using equations (1) and (2). However, unlike in CBSEM, there is not necessarily need to distinguish between endogenous and exogenous variables in PLS when considering model equations (Henseler et al., 2008). Thus, if considering all variables as endogenous, the (reflective) outer model can be written more simply as

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon} \quad , \tag{4}$$

and the inner model as

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \quad . \tag{5}$$

The PLS algorithm is essentially a sequence of regressions in terms of weight vectors (Henseler et al., 2008). In contrast to CBSEM, the analysis starts by estimating a proxy value for each latent construct as a summated scale of its indicator variables. These proxy values are then used to run regression models for each latent variable in the model. Then new proxy values are calculated based on the results of this inside approximation and are used in regressing the latent constructs on each of their indicators. The results of this outside approximation are used to calculate weights for each indicator-latent variable -relationship after which a new set of proxy values are used as a starting point for new round of inside approximation. These two steps are repeated until the change in outer weights between two iterations drops below a predefined limit (for more details, see Chin, 1998a).

4. Controlling for common method variance

4.1. Common method variance

Common method bias is a subset of method bias (Burton-Jones, 2009). It arises in quantitative research when the covariance caused by the measurement approach rather than the measured trait causes measured relationships between two constructs to either inflate or attenuate compared to the true value (Williams & Brown, 1994). This is a frequently encountered problem especially with survey studies. Classical test theory (cf. Nunnally, 1967) that provides the theoretical foundations for much of the measurement tools that IS researchers use assumes that each person or organization measured has a true score and that any measured score is a function of this true score and measurement error:

$$\boldsymbol{X}_i = \boldsymbol{T}_i + \boldsymbol{S}_i + \boldsymbol{e} \quad , \tag{6}$$

where $\boldsymbol{X}_i$ is a vector containing the measured scores in item i, $\boldsymbol{T}_i$ is a vector of the true scores, $\boldsymbol{S}_i$ a vector of item specific but reliable error components, and $\boldsymbol{e}$ is a vector of random errors. A problem with common method variance arises when the item specific components of the measured scores correlate across items. In general, these unwanted measurement effects can cause bias in the statistical analyses if present in the data and not properly controlled. A key problem with common method bias is that these effects that cause loss of construct validity are sometimes difficult to detect and are often not detected with standard tests for discriminant and convergent validity (Straub, Boudreau, & Gefen, 2004; Richardson et al., 2009). The problem of identifying method variance is complicated because various sources of error variance can coexist and overlap. Additionally, the variance resulting from method can be congeneric or noncongeneric, that is, affecting each item either equally or differently (Richardson et al., 2009).

While it is generally agreed on that measurement results are affected by both the measurement approach and the measured trait, opinions differ on how commonly the variance caused by measurement approach causes significant bias in the results. In addition, the research results relating to the existence and significance of common method bias remain mixed (Richardson et al., 2009). While the evidence on the overall impact of common method variance remains inconclusive, scholars generally agree that common method variance can cause problems. However, only a minority of IS studies explicitly address these concerns: For example, in their review, Woszczynski and Whitman (2004) observed that only 12 of the reviewed 116 articles with potential common method problem explicitly noted it and even fewer attempted to control for it. While controlling the effects of common method variance can be done on multiple levels starting from the study design and data collection, this study focuses only on statistical remedies available after the data has been collected.

For examining the extent to which common method variance is present in the data, most commonly used method is Harman's single factor test (cf. P. M. Podsakoff & Organ, 1986). In this technique exploratory factor analysis is utilized to evaluate the amount of variance in observed variables that can be explained by a single factor. This is determined by examining the first factor of the unrotated factor solution. If either a single strong factor emerges or the first factor loads significantly on all items, common method variance is most likely present in the data (P. M. Podsakoff & Organ, 1986). However, there are three potential problems with this technique: it is very unreliable (cf. Kemery & Dunlap 1986), no clear guidelines are available as to when this technique indicate problematic amount of method variance, and it does nothing to actually control the method variance. Besides Harman's test, another method has recently appeared in the toolbox of IS researchers, called the marker variable technique (Lindell & Whitney, 2001). With this technique, a researcher includes a priori defined marker variable that should be theoretically unrelated to the study variables and then calculates the correlation between this variable and the study variables. Since the variables are not theoretically related, the correlation is assumed to solely result from method variance and can be partialed out from other correlations in the study.

Besides these general methods, several techniques have been developed for structural equation modeling for explicitly modeling the common method variance in the models.

These techniques are generally developed for CBSEM. Two most commonly used methods include various method factor design and correlated uniqueness designs. Podsakoff and his colleagues (2003) provide a good overview of these methods: Generally the idea is to add error correlations or factors to the analysis thus allowing the covariance that results from the measurement to escape from the model rather than affect the substantive regression or correlation relationships. Technically these methods rely on partialing out the reliable error variance on the indicator level so that it does not affect the parameter estimates in the structural part of the model. The method factor designs can further be classified into two groups depending on whether only the indicators of the study variables are used in the analysis or whether extra marker variables are included as indicators for the method factor. While particularly these designs have been adopted in previous studies (e.g. Ye, Marinova, & Singh, 2007; Alge, Ballinger, Tangirala, & Oakley, 2006; Agustin & Singh, 2005), the effectiveness of the method factor design has only recently been tested in a simulation setting: In their study, Richardson and his colleagues (2009) provided evidence that the method factor design with marker variables, the CFA marker technique, could in most cases effectively reduce the bias caused by method variance even if the marker variables were non-ideal by correlating with the study variables. In their analysis, the unmeasured latent method construct approach often produced less accurate results.

4.2. Implementation to PLS path modeling

While several techniques exist for controlling common method variance in CBSEM, they are generally not directly applicable to PLS path modeling. The reason for this is that these techniques rely on partialing the variance to model variance and error variance on an indicator level, but with PLS the indicators are only weighted and summed without partialing variance. For example, if one third of the variance of each indicator is congeneric measurement variance, the total variance of the latent variables will consists of one third method variance regardless of how the indicators are weighted. While introducing a method factor to a PLS model might provide an estimate of the strength of the method variance, this approach does not prevent covariance caused by the measurement approach affecting the inner model. While PLS path modeling has been shown to be sensitive to various sources of method variance (A. Schwarz, C. Schwarz, & Rizzuto, 2008), no methods for controlling for common method variance have been developed in the PLS context. This far the users of PLS path modeling have only had the option to partial out correlations between marker variable from the study correlations and then use this corrected data for the main analyses (Lindell & Whitney, 2001).

The approach proposed in this study for controlling for method variance is to control it during the inner estimation. A somewhat similar approach, although implicitly and only for diagnostic purposes, has been previously adopted in recent PLS papers (e.g. Shutao Dong et al. 2009; Pavlou et al. 2007). In these papers, a proxy for common method variance was formed by conducting an explanatory factor analysis on all items in the model, and using the first emerging unrotated factor as a control variable in the inner model. Thus this approach is similar to using Harman's single factor test in obtaining a proxy for common method variance. Similarly to these two mentioned papers, this study suggests that a method factor is included as a predictor for all endogenous latent constructs in the model. Thus the common method variance is controlled for in the inner model rather than partialing it out during the outer estimation. Conceptually, this would mean that the calculated values from the outer estimation for the latent constructs are a result of the true relationships between the constructs and error variance caused by measurement. This approach differs from the previously suggested ways by using a directly measured method factor rather than building a proxy based on the substantive items in the model. The indicators of the method factor should be theoretically unrelated to any of the constructs of interest and preferably not correlated except for the correlation caused by sharing the same method. Contrary to CBSEM based CFA marker approach, the method factor should not load on the indicators of the study constructs due to the fact that a construct loading on the items of another independent variable would be severely collinear and shared items with the dependent variable would cause the coefficients between method factor and endogenous constructs to be inflated. As a joint effect, this would severely bias the estimates of the coefficients between the actual model constructs. Figure 1 illustrates the proposed approach.

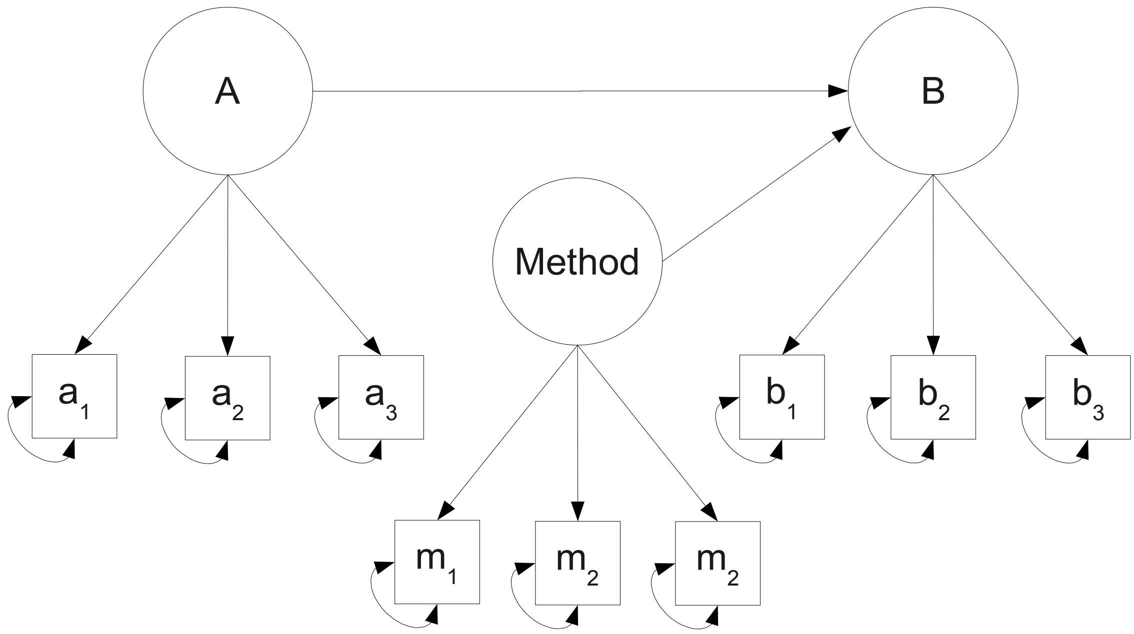

Figure 1: Simplified example of the proposed approach

5. Simulation

The proposed approach for controlling common method variance with PLS was tested by using Monte Carlo simulation. Several PLS and CBSEM models with and without measured method factor were compared under varying conditions of sample size, level of method variance, and number of method indicators. Similar approach has been previously used when testing new approaches for PLS modeling (Qureshi & Compeau, 2009; Chin et al., 2003) as well as more recently when testing the effectiveness of different approaches for controlling common method variance (Richardson et al., 2009).

5.1. Monte Carlo simulation

Use of Monte Carlo simulation has become common when inspecting properties related to structural equation modeling. Analytical statistical theory can address some research questions, but finite sample properties of SEM estimators are often beyond the reach of the established asymptotic theory (Paxton, Curran, Bollen, Kirby, & Chen, 2001). In some cases the distributions are not known even asymptotically (e.g., several fit measures).

Monte Carlo simulation concerns with using simulated random numbers in examining the properties of the distributions of random variables (Paxton et al., 2001). It allows 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 evaluation of the behavior of a statistic.

When performing Monte Carlo simulation in SEM context, the researcher first creates a model with known population parameters (i.e., the values are set by the researcher) (Paxton et al., 2001). 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 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.

5.2. Simulation design

Simulation design was carried out by following the guidelines set by Paxton, Curran, Bollen, Kirby, and Chen (2001). As stated above, the first step is to determine the population model for data generation. Typically four aspects should be considered when choosing the model (Hancock & Mueller, 2006): model type, model size, model complexity, and model parameters. Based on these aspects, four criteria were set for an appropriate population model: First, the model must have both measurement and structural part. Thus, for example, a CFA model would be inappropriate because it only includes the measurement model. Second, the model needs to have a level of size and complexity that is frequently met in practice. The purpose of this criteria was to reflect the practical situations in IS as well as possible, thus increasing generalizability of these results. In addition, a very simple model would decrease the validity of the results, whereas a too complex model would set high requirements for computing power and would also make interpretation of the results difficult. Third, there must be a decent amount of previous research studies on the model. This criteria was necessary in order to set realistic and appropriate coefficient values for the population model before data generation.

Based on the above-mentioned criteria, the technology acceptance model (TAM) was chosen as the structural part of the population model. This model has been used in several recent studies (Sharma, Yetton, & Crawford, 2009; Malhotra, Kim, & Patil, 2006). In order to meet the requirement of appropriate model complexity, the extended TAM model was chosen instead of the original TAM model. This extended model is presented in a meta-analysis paper by Schepers and Wetzels (2007) who also estimated the standardized structural coefficients for TAM on the basis of the correlation matrix obtained from an aggregation of several individual research studies. The model and the parameter values are presented in Figure 2. Correlations between exogenous latent variables were set to zero for two reasons: First, by allowing latent variable effects only through the regression paths the model remained more parsimonious. Second, the appropriate level of correlation would be difficult to determine, as it varies between different research settings. Altogether, the chosen model consists of six latent variables and ten regression paths between them as illustrated in Figure 2.

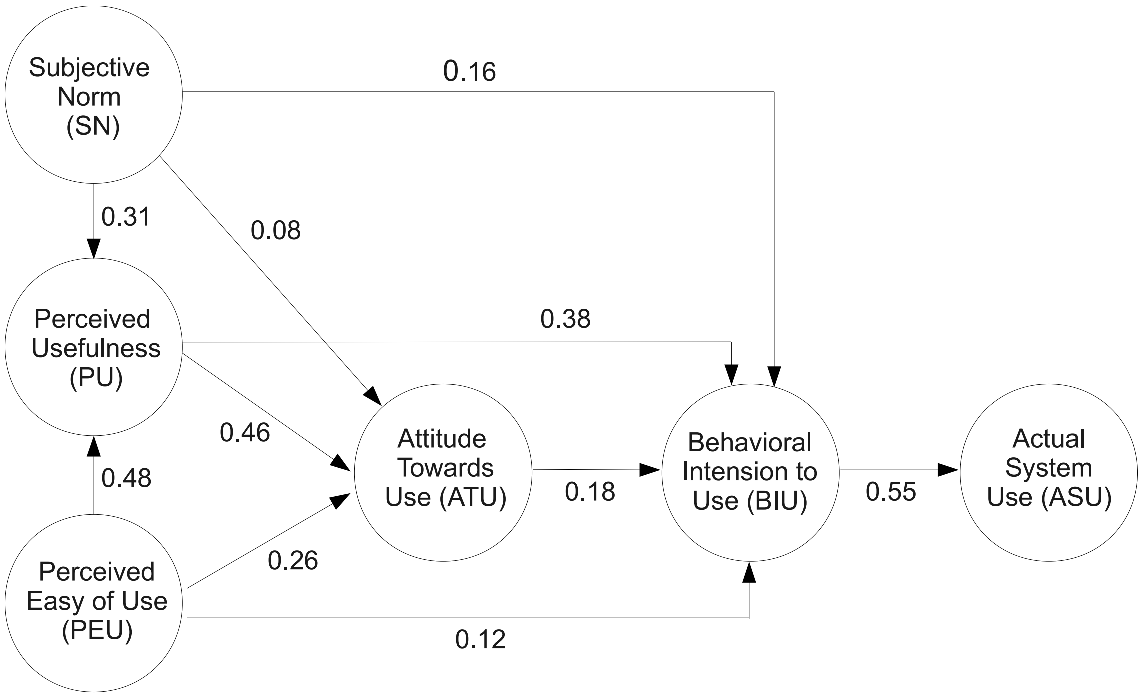

Figure 2: Structural part of the population model used in simulation

The next task after choosing the appropriate structural part of the population model was to define properties of the measurement model. The first question was whether to use reflective or formative indicators. Reflective indicators were chosen for three reasons: First, since the purpose was to compare the results of analysing the same set of data with PLS and CBSEM, difficulties in using formative indicators in CBSEM analysis was a concern (Chin, 1998b) and using different modes of measurement would potentially bias the results (MacKenzie, P. M. Podsakoff, & Jarvis, 2005). Second, survey research that is most commonly used to test TAM model predominantly uses reflective measurement. Third, formative measurement has been criticized lately in the methodological literature (Howell, Breivik, & Wilcox, 2007).

The amount of indicators was set to three for all latent variables in the model. This number was chosen for three reasons: First, some recent works have emphasized that three indicators per construct in confirmatory analyses is an optimal number (Little et al. 1999) [1]. Second, three indicators is a fairly common amount in IS research (Chin et al., 2003). Third, it is also low enough to reduce the potential convergence issues due to over-identification of the CBSEM models that are fitted to the data. By choosing three indicators per latent variable, the total amount of indicator variables was 18 for the underlying TAM model.

As the purpose was to inspect the effects of common method variance, a factor representing the source of variance due to common method was added to the population model. An additional latent variable was included for this purpose. This latent method factor was set to be uncorrelated with all the other latent variables, but had a loading on each of the indicator variables in the model. These loadings were constrained to be equal in order to model congeneric method variance. With this approach, the latent method factor produced equal, systematic, variance in all of the indicators, thus simulating common method variance.

The proposed approach concerns using marker variables in controlling common method variance, and thus a set of marker variables was created. These indicators were designed to reflect the common method variance in the model, and they were set uncorrelated with all the other indicators in the model except for the correlation caused by the method factor. 18 such items were created to enable modeling the impact of varying number of marker variables in the model. For survey research, this many extra indicators are not uncommon, since often only a subset of data from a larger survey are used in analyses for one paper.

The indicator variables were set to be centered, and variances of all latent and indicator variables were set to one. The factor loadings between the indicators and the model

1 There is also research that suggests that "more is better" when considering the amount of indicators per latent variables. See for example Marsh et al. (1998)

constructs were set to 0.70, since this value has been recommended as a minimum standard in the IS literature and has been consequently used in prior Monte Carlo study by Chin, Marcolin, and Newsted (2003). The loadings of indicators on the method factor were included as a varying condition in the experimental design. These loadings were, however, restricted to be equal with each other, thus reflecting the similar effect of measurement method for each item.

After the population model for data generation was chosen, the next step was to determine the conditions to vary in the simulation. In total, four varying conditions were chosen: sample size, level of common method variance, amount of marker variables, and the fitted model.

The first varying condition was sample size. In total, four sample sizes were used: 125, 250, 500, and 1000. The initial purpose was to set the smallest size close to 100, as it is often considered as an important threshold (Paxton et al., 2001). However, the size of the model and the decision to use CBSEM as a comparison technique prevented using this small sample size since the CBSEM coefficient estimates cannot be determined if the amount of estimated parameters exceeds the amount of observations. Due to this, the minimum sample size was set to 125.

The second condition was to vary the level of common method variance in the population model. Four different levels for common method variance were decided: *None*, *Little*, *Moderate*, and *High*. The level of common method variance was controlled in data generation by varying the factor loadings of the indicators on the method factor. The respective values were set as 0.001, 0.1, 0.3, and 0.5. The reason for setting the loading for *None* level as 0.001 instead of 0 was that choosing a non-zero value simplified the data generation and analysis, and the difference is yet insignificant in practice.

The third condition to vary was the amount of marker variables. The number of method indicators was included as a varying condition since the random intercorrelations between marker variables reduce the ability of these items to accurately reflect method variance and increasing the number of markers was a simple way to reduce the effect of each individual random correlation between the items. Moreover, four levels for comparison were chosen: 3, 6, 9, and 18. These were chosen to reflect the amount of indicators in the underlying model. Three indicators per latent variable were chosen for data generation, and thus three was chosen as the first method indicator amount. The total amount of indicators in the model was 18, and thus it was chosen as the upper limit for method indicators. Amounts of 6 and 9 were added from between the above decided values in order to estimate the necessary amount of indicators that is sufficient for the suggested method to function correctly.

The fourth varying condition was the fitted model. In total, four different models were used: First, a PLS model was included that was formulated according to the population model with the exception of omitting the common method factor. Thus this model was misspecified for analysing data with common method variance. The purpose of this model was to act as a control to enable evaluating the effect of common method variance on parameter estimates in an uncontrolled model, and thus to function as a reference point when evaluating the efficiency of the suggested approach. Second, a PLS model was chosen with modification to account for common method variance. As described in the previous section, this model includes an additional latent construct utilizing only method indicators that are theoretically not related with other indicators in the model. This construct was added as a predictor for endogenous variables in the inner model to control for common method variance. The results of this model were compared to the misspecified PLS model described above. Third, a maximum likelihood estimated CBSEM model was added. This model is formulated similarly to the first, misspecified PLS model, that is, without correction for the method variance. The purpose of this CBSEM model was primarily to provide a comparison point for the PLS models to asses the well-know feature of PLS to overestimate factor loadings and underestimate latent path coefficients (Chin et al., 2003). The fourth model is a CBSEM model with correction for common method variance with measured latent method factor. In this technique, items are allowed to load on their theoretical constructs, as well as on a latent method factor that has also its own method indicators reflecting the presumed cause of the method bias. In the present study, this model was included as a control model that effectively controls for method variance (Richardson et al., 2009) further enabling us to asses the merits of the proposed PLS based approach. Figure 3 and Figure 4 show the CBSEM and PLS models where corrections for common method variance are utilized.

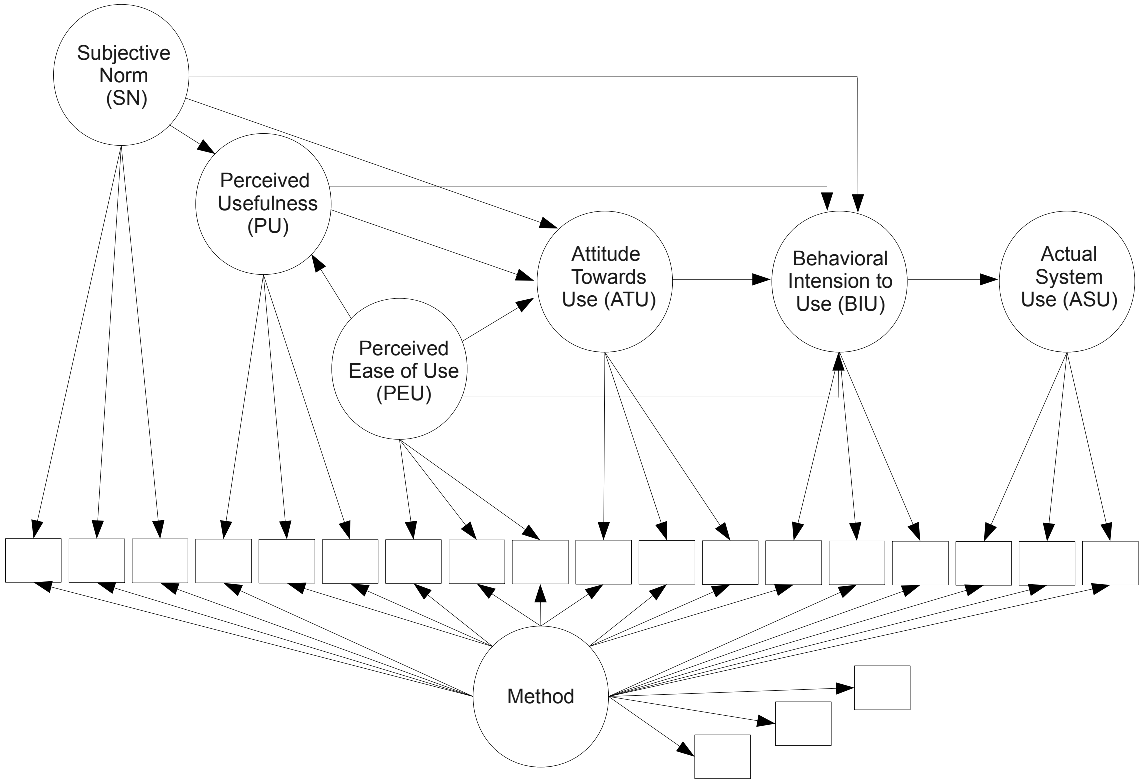

Figure 3: CFA marker variable with CBSEM

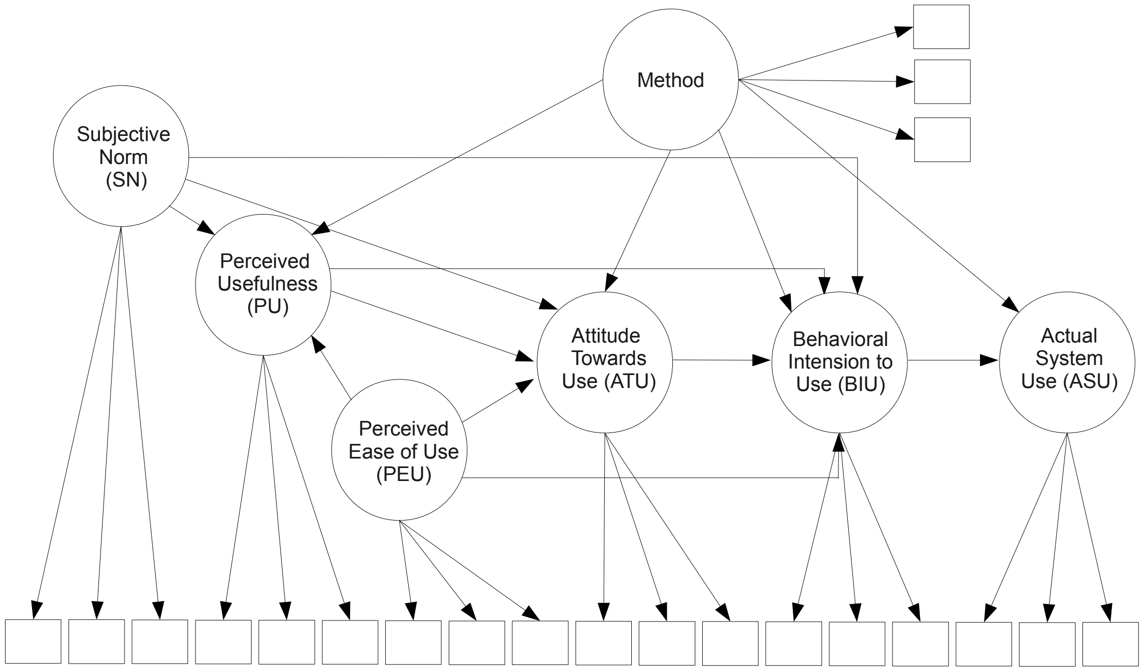

Figure 4: Latent marker variable with PLS

In total, the combination of these four factors resulted in 256 (4 x 4 x 4 x 4) unique modeling conditions. The amount of replications for each condition was set to 500 as it is often used in SEM Monte Carlo studies (Hancock & Mueller, 2006). As there are possibly non-converged or improper solutions (such as including negative variances) when estimating CBSEM models, the suggestions of Paxton et al. (2001) was utilized to use the correctly specified CBSEM model to evaluate the quality of each replication and thus replications were generated until 500 "good" replications were obtained.

In total, four data sets of each containing 500 separate replications were generated, one for each of the four chosen levels of common method variance. Different sample sizes were analysed by limiting the amount of observations from each sample. The amount of method indicators was set to the maximum level for data generation, that is, to 18. Since the method indicators were not correlated with the other indicators except for method variance, the amount of these items did not affect any other items in the data generation. Only the necessary subset of these items were used for for each modeling condition. In order to achieve convergence as quickly as possible for each replication, the population parameters were used as starting values when estimating the CBSEM models, as suggested by Paxton et al. (2001).

Data were generated from multivariate normal distribution by using Mplus 5.1 structural equation modeling software. The same software was used for estimation of the CBSEM models. For PLS modeling, version 0.1-4 of plspm-package of the R statistical software environment was used. This software was chosen instead of some of the more popular graphical tools because it is one of the very few PLS modeling tools that are currently available that provide support for command line usage, which was required for efficient Monte Carlo simulation.

6. Results

Before using any explicit controls for common method variance, its presence should be tested (Richardson et al., 2009). As Harman's single factor test is currently the most popular test for detecting common method variance, it was applied to the generated data before proceeding to the main analyses. Although Harman's single factor test is widely used, some confusion seems to exist in which factor extraction method should be used. Two different methods are generally used: Principal axis factoring (PAF) and principal component factoring (PCF). The former of these two considers only the variance that is shared between items and ignores the variance considered as random error. As the purpose of Harman's single factor method is to identify the amount of reliable error variance that is correlated between items, PAF seems more appropriate. However, PCF is also used by some researchers (e.g. Pavlou & El Sawy, 2006). Therefore results are reported by using both of these methods.

The results of Harman's single factor test for different data are presented in Table 1. As stated earlier in section 4.1. some researchers have found this method to be unreliable, and the results here corroborate these findings. The figures in the table are percentages describing the amount of variance explained by the first unrotated factor. The first factor tends to explain over half of the variance even in the case where common method variance does not exist in reality. Another observation is that the differences in figures between different common method variance levels are relatively small. Thus Harman's test clearly does not provide reliable estimates for the level of common method variance in this case. A potential explanation for the apparently overestimated amount of common variance in data is that no error correlations were included in data generation. Heterogeneous, pairwise error correlations often encountered in practice would decrease the amount of variance shared by all of the variables in the model, and thus reduce the eigenvalue of the first factor in Harman's test. Another reason for these results is that no correlations were allowed between exogenous latent variables in data generation due to reasons explained in the previous chapter. Adding this correlation would potentially reduce the amount of common variance through the same mechanism as adding pairwise error correlations to the model.

Table 1: Harman's single factor tests for different data

	Number of observations							
	125		250		500		1000	
Method variance	PAF	PCF	PAF	PCF	PAF	PCF	PAF	PCF
none	52.2	21.3	58.3	20.8	63.3	20.7	66.5	20.6
Little	52.5	21.7	60.5	21.5	66.1	21.4	69.3	21.4
Medium	58.5	25.2	65.2	24.5	70.7	25.0	73.4	25.1
Large	65.5	31.0	73.3	31.0	77.8	31.0	80.4	31.2

The next step was to fit the four models – correctly specified and misspecified PLS and CBSEM models – to the generated data. These models were fitted to the each data set with different sample size, level of common method variance, and number of marker variables. Table 2 presents the results for the case where sample size was 250, loadings on the method factor were 0.3, and 9 marker variables were used. The results provide several interesting observations: First, CBSEM models can generally restore the original indicator loadings relatively well, whereas PLS models tend to systematically overestimate these. The substantive indicator loadings were set to 0.7 in data generation, but PLS models tend to restore values closer to 0.8. These differences also turned out to be statistically significant. However, this tendency of PLS to overestimate indicator loadings is a known feature of PLS (Chin et al., 2003).

Table 2: Results of model fitting when N=250, common method variance level is 0.3, and 9 method indicators are used

	PLS path model				Structural equation model				True
	Correct		Misspesified		Correct		Misspesified		population
Path	Mean	SD	Mean	SD	Mean	SD	Mean	SD	value
0 ASU ON BIU	0.410	0.052	0.438	0.049	0.547	0.127	0.603	0.112	0.550
0 ATU ON PEU	0.191	0.058	0.205	0.058	0.277	0.162	0.331	0.158	0.260
0 ATU ON PU	0.310	0.060	0.322	0.059	0.457	0.151	0.462	0.149	0.460
0 ATU ON SN	0.084	0.057	0.099	0.056	0.083	0.149	0.139	0.131	0.080
0 BIU ON ATU	0.175	0.063	0.182	0.062	0.180	0.142	0.187	0.128	0.180
0 BIU ON PEU	0.105	0.061	0.116	0.060	0.116	0.174	0.180	0.162	0.120
0 BIU ON PU	0.256	0.061	0.264	0.060	0.390	0.163	0.391	0.157	0.380
0 BIU ON SN	0.117	0.054	0.129	0.054	0.158	0.153	0.224	0.129	0.160
0 PU ON PEU	0.287	0.058	0.311	0.057	0.489	0.157	0.588	0.146	0.480
0 PU ON SN	0.194	0.057	0.218	0.056	0.309	0.139	0.405	0.124	0.310
1 ASU BY ASU1	0.792	0.038	0.792	0.039	0.686	0.093	0.696	0.084	0.700
1 ASU BY ASU2	0.790	0.040	0.790	0.040	0.690	0.097	0.696	0.086	0.700
1 ASU BY ASU3	0.792	0.040	0.792	0.041	0.693	0.097	0.699	0.087	0.700
1 ATU BY ATU1	0.794	0.039	0.794	0.039	0.680	0.098	0.686	0.085	0.700
1 ATU BY ATU2	0.792	0.040	0.792	0.040	0.672	0.095	0.683	0.086	0.700
1 ATU BY ATU3	0.793	0.038	0.793	0.038	0.674	0.092	0.683	0.082	0.700
1 BIU BY BIU1	0.795	0.037	0.795	0.036	0.675	0.101	0.674	0.080	0.700
1 BIU BY BIU2	0.795	0.037	0.795	0.037	0.672	0.096	0.674	0.080	0.700
1 BIU BY BIU3	0.793	0.035	0.793	0.035	0.673	0.092	0.671	0.077	0.700
1 PEU BY PEU1	0.755	0.054	0.755	0.054	0.695	0.115	0.757	0.095	0.700
1 PEU BY PEU2	0.755	0.053	0.755	0.053	0.692	0.118	0.755	0.096	0.700
1 PEU BY PEU3	0.757	0.050	0.757	0.050	0.691	0.118	0.758	0.092	0.700
1 PU BY PU1	0.782	0.038	0.782	0.038	0.681	0.100	0.686	0.084	0.700
1 PU BY PU2	0.786	0.038	0.786	0.038	0.687	0.102	0.692	0.088	0.700
1 PU BY PU3	0.783	0.041	0.783	0.041	0.683	0.097	0.689	0.086	0.700
1 SN BY SN1	0.756	0.061	0.756	0.061	0.692	0.123	0.757	0.094	0.700
1 SN BY SN2	0.750	0.070	0.750	0.070	0.692	0.126	0.759	0.101	0.700
1 SN BY SN3	0.755	0.068	0.755	0.068	0.699	0.120	0.761	0.102	0.700
2 ASU ON METHOD	0.123	0.060							
2 ATU ON METHOD	0.087	0.062							
2 BIU ON METHOD	0.078	0.058							
2 PU ON METHOD	0.121	0.059							
3 METHOD BY ASU1					0.302	0.132			0.300
3 METHOD BY ASU2					0.298	0.136			0.300
3 METHOD BY ASU3					0.299	0.134			0.300
3 METHOD BY ATU1					0.298	0.136			0.300
3 METHOD BY ATU2					0.308	0.137			0.300
3 METHOD BY ATU3					0.303	0.144			0.300
3 METHOD BY BIU1					0.306	0.139			0.300
3 METHOD BY BIU2					0.310	0.133			0.300
3 METHOD BY BIU3					0.299	0.132			0.300
3 METHOD BY PEU1					0.295	0.123			0.300
3 METHOD BY PEU2					0.293	0.121			0.300
3 METHOD BY PEU3					0.299	0.126			0.300
3 METHOD BY PU1					0.301	0.133			0.300
3 METHOD BY PU2					0.309	0.134			0.300
3 METHOD BY PU3					0.304	0.137			0.300
3 METHOD BY SN1					0.301	0.120			0.300
3 METHOD BY SN2					0.301	0.122			0.300
3 METHOD BY SN3					0.297	0.116			0.300
4 METHOD BY M1	0.395	0.140			0.288	0.099			0.300
4 METHOD BY M2	0.398	0.138			0.291	0.099			0.300
4 METHOD BY M3	0.401	0.142			0.293	0.103			0.300
4 METHOD BY M4	0.392	0.149			0.287	0.097			0.300
4 METHOD BY M5	0.399	0.144			0.294	0.103			0.300
4 METHOD BY M6	0.392	0.135			0.287	0.100			0.300
4 METHOD BY M7	0.405	0.142			0.297	0.101			0.300
4 METHOD BY M8	0.401	0.136			0.292	0.098			0.300
4 METHOD BY M9	0.404	0.140			0.299	0.097			0.300

In addition to overly large indicator loadings, PLS also tends to overestimate the loadings of the method indicators on the method factor. The population values were 0.3 for all of these loadings, but PLS tends to systemically restore values closer to 0.4.

The third observation from Table 2 is related to the regression coefficients restored by different models. The correctly specified CBSEM model was generally able to restore the correct coefficients, whereas misspecified CBSEM model had a tendency to overestimate them. Both PLS models, on the other hand, tended to underestimate the coefficients. The deviation from the correct values was worse for the misspecified model whereas the correctly specified model provided slightly better estimates. This underestimation of the regression coefficients is, however, a previously recognized feature of PLS. Thus it is more interesting that the correctly specified PLS model was able to provide lower estimates than the misspecified PLS model. Since the previous research shows that PLS modeling is not immune to common method bias (A. Schwarz et al., 2008), this implies that the proposed approach provide results that are less biased than the results without controlling for common method variance.

In the previously presented table only one level of common method variance was used. To examine how well the proposed approach is able to account for common method variance, mean estimates for coefficients were calculated and compared across all levels of common method variance. Table 3 illustrates the results in the case where the sample size was 1000 and 9 marker variables were used. At the smallest level of common method variance – that is, when none is present – the estimates does not significantly differ between correct and misspecified models. However, as the level of common method variance increases, the estimates in the misspecified model become inflated considerably faster than in the correct model. Neither model is totally immune to the common method variance, but the correct model seems to be able to scale down its effect to some extent. The factor loadings, however, are not affected by the correction, and are thus overestimated. In other words, the proposed approach can largely control for the potential violations of internal validity caused by common method variance but does not address concerns of estimates of construct validity being inflated. Another interesting observation is that the loadings on the method factor confine the amount of common method variance in data relatively well. Thus they can potentially provide a useful diagnostic test for the extent to which common method variance is present in data.

Table 3: Average PLS estimates for different levels of common method variance, when N=1000 and 9 method indicators are used

	Mean PLS estimates for different levels of method variance							
	Correct model				Misspecified model			
Path	0	.1	.3	.5	0	.1	.3	.5
0 ASU ON BIU	0.381	0.386	0.411	0.423	0.386	0.393	0.438	0.504
0 ATU ON PEU	0.172	0.178	0.191	0.204	0.174	0.180	0.205	0.238
0 ATU ON PU	0.305	0.305	0.312	0.320	0.310	0.310	0.323	0.342
0 ATU ON SN	0.064	0.073	0.084	0.095	0.065	0.074	0.098	0.131
0 BIU ON ATU	0.168	0.163	0.175	0.177	0.171	0.166	0.182	0.191
0 BIU ON PEU	0.095	0.098	0.105	0.113	0.096	0.100	0.115	0.138
0 BIU ON PU	0.248	0.253	0.258	0.266	0.252	0.258	0.265	0.279
0 BIU ON SN	0.107	0.108	0.116	0.126	0.108	0.110	0.128	0.154
0 PU ON PEU	0.260	0.267	0.287	0.306	0.266	0.272	0.311	0.363
0 PU ON SN	0.172	0.179	0.194	0.213	0.175	0.182	0.217	0.270
1 ASU BY ASU1	0.780	0.781	0.792	0.810	0.779	0.780	0.791	0.810
1 ASU BY ASU2	0.776	0.778	0.791	0.811	0.776	0.778	0.791	0.811
1 ASU BY ASU3	0.777	0.779	0.793	0.809	0.777	0.780	0.793	0.809
1 ATU BY ATU1	0.781	0.782	0.794	0.814	0.781	0.782	0.794	0.814
1 ATU BY ATU2	0.781	0.785	0.793	0.815	0.781	0.785	0.793	0.815
1 ATU BY ATU3	0.783	0.783	0.794	0.813	0.783	0.783	0.794	0.813
1 BIU BY BIU1	0.782	0.784	0.795	0.813	0.782	0.784	0.795	0.813
1 BIU BY BIU2	0.782	0.785	0.795	0.815	0.782	0.785	0.795	0.815
1 BIU BY BIU3	0.784	0.784	0.794	0.814	0.784	0.784	0.794	0.814
1 PEU BY PEU1	0.738	0.738	0.756	0.785	0.738	0.738	0.756	0.785
1 PEU BY PEU2	0.738	0.740	0.755	0.783	0.738	0.740	0.755	0.783
1 PEU BY PEU3	0.735	0.743	0.758	0.784	0.735	0.743	0.758	0.784
1 PU BY PU1	0.771	0.772	0.782	0.806	0.771	0.773	0.782	0.806
1 PU BY PU2	0.768	0.766	0.785	0.802	0.768	0.765	0.785	0.802
1 PU BY PU3	0.769	0.774	0.783	0.805	0.769	0.774	0.783	0.805
1 SN BY SN1	0.727	0.726	0.754	0.782	0.727	0.726	0.754	0.782
1 SN BY SN2	0.726	0.731	0.750	0.781	0.727	0.731	0.750	0.781
1 SN BY SN3	0.729	0.734	0.756	0.784	0.730	0.734	0.756	0.784
2 ASU ON METHOD	-0.007	0.027	0.122	0.218				
2 ATU ON METHOD	0.004	0.023	0.086	0.129				
2 BIU ON METHOD	-0.008	0.029	0.075	0.107				
2 PU ON METHOD	-0.001	0.038	0.119	0.176				
4 METHOD BY M1	0.080	0.109	0.399	0.529				
4 METHOD BY M2	0.086	0.122	0.401	0.533				
4 METHOD BY M3	0.081	0.138	0.401	0.532				
4 METHOD BY M4	0.089	0.119	0.393	0.535				
4 METHOD BY M5	0.074	0.128	0.400	0.532				
4 METHOD BY M6	0.084	0.128	0.398	0.533				
4 METHOD BY M7	0.093	0.130	0.405	0.535				
4 METHOD BY M8	0.078	0.129	0.401	0.533				
4 METHOD BY M9	0.084	0.136	0.407	0.534				

As the proposed approach seems to be able to scale down the bias caused by common method variance, the next question is how many method indicators is necessary to include in the model. This feature was examined by calculating average bias in regression coefficients in several sample sizes, levels of common method variance, and numbers of method indicators. These results are presented in Table 4. The figures in this table are mean percentages of inflation of latent regression coefficients when comparing the focal modeling condition to results that were produced with a model without controlling for method variance using data uncontaminated with common method variance. These results provide several observations: First, if no common method variance is present, applying remedies for it tend to bias results downward. This finding in line with previous studies (Richardson et al., 2009) and the key question is whether this bias is sufficient to cause the interpretation of the results to change. The bias seems to depend on both sample size and the amount of utilized method indicators: adding more method indicators increases the bias, whereas increasing the sample size tensd to decrease it.

Another observation is that if common method variance is present in data and no controls are added, the coefficients tend to become upward biased resulting in Type I error. The severity of this bias tend to increase as the level of common method variance in data increases. When method controls are added to the data, the coefficients generally become less inflated. Especially with high levels of common method variance, adding more method indicators tend to decrease the bias considerably. The largest amount of method indicators tested was 18, and up to that point the bias tended to scale down when more method indicators were added.

Table 4: Bias caused by presence of common method variance when varying sample size, number of method indicators, and population common method variance

	Mean bias for different number of method indicators					Mean SD of path estimate for different number of method indicators				
	none	3	6	9	18	none	3	6	9	18
No method variance										
125	0.0	-0.7	-1.9	-3.2	-7.3	0.0	0.7	1.2	1.2	1.8
250	-0.0	-0.3	-1.0	-1.6	-3.9	0.0	0.4	0.3	0.4	0.9
500	0.0	-0.2	-0.5	-0.9	-2.0	-0.0	0.0	0.0	-0.0	0.1
1000	0.0	-0.1	-0.3	-0.4	-1.0	-0.0	0.1	0.1	0.1	0.1
Little method variance										
125	2.5	1.7	0.4	-1.0	-5.0	1.5	1.7	1.9	2.1	2.2
250	2.2	1.8	1.1	0.4	-1.8	-1.0	-0.9	-0.6	-0.6	-0.5
500	1.8	1.6	1.2	0.8	-0.5	1.4	1.4	1.4	1.5	1.5
1000	1.9	1.8	1.5	1.3	0.6	1.8	1.9	2.0	2.1	2.1
Medium method variance										
125	13.4	9.9	6.6	4.0	-1.4	-2.6	-1.4	-1.3	-0.7	0.3
250	13.8	10.8	8.2	6.2	2.3	-4.1	-3.3	-3.1	-2.9	-2.1
500	14.1	11.4	9.2	7.6	4.5	-1.5	-1.0	-0.8	-0.5	0.4
1000	14.5	11.9	9.9	8.5	5.7	-0.8	-0.3	-0.1	0.1	0.5
Large method variance										
125	29.6	19.7	13.7	10.2	4.5	-5.4	-3.0	-2.1	-1.1	-0.2
250	30.3	20.4	15.1	11.9	6.6	-4.5	-1.9	-1.1	-0.5	0.3
500	30.4	20.6	15.4	12.4	7.4	-2.6	0.1	1.2	1.9	2.4
1000	31.1	21.4	16.4	13.3	8.5	-3.2	-0.8	0.5	1.2	2.0

7. Discussion and conclusions

Common method variance is a commonly encountered source of bias especially when analysing survey studies, and is continuously provoking attention among IS researchers. In structural equation modeling several tools exist for controlling method bias, but most of them are developed for CBSEM and are not directly applicable to other approaches. This study introduces a method for controlling common method variance in PLS path modeling context. The proposed approach was tested under several conditions using Monte Carlo simulation. The results suggest that the proposed approach can significantly decrease the bias caused by common method variance in estimates when using PLS. The extent to which this approach is useful depends on the level of common method variance present in data. If the suggested remedies are added when common method variance does not exist in reality, the results tend to become biased. Therefore it is recommended to diagnose the extent to which common method variance is present before utilizing this approach and always running the model with and without the method factor.

The amount of marker variables included should reflect the sample size and complexity of the data: for large sample sizes and large method variance levels, more indicators seem to produce better results. In some occasions, it is even reasonable to include as many marker variables as there are substantive variables in the model. When using the proposed approach, it is recommended to use marker variables that are correlated only due to method effect. However, as this condition is seldom fulfilled in practice, it is recommended to verify that no high correlations exist among marker variables, and if necessary, exclude highly correlating variables from the data.

Although the presented approach for controlling common method variance using PLS path modeling clearly shows positive results, there are some limitations: First of all, no error correlations were included in data generation. These correlations often exist in practice, and thus this approach should also be tested using data including, for example, small, random, pairwise correlations between indicator variables. Second, exogenous latent variables were considered to be uncorrelated with each other in data generation, although this situation seldom exist in practice. Third, in this analysis, marker variables were "ideal" in a sense that they were correlated with all other variables in the model only through method effect. Fourth, only "good" samples were used. This means that only samples were used that did converge when using correctly specified CBSEM models. Although this procedure is often used in SEM Monte Carlo studies, this means that the results of this study can generalized only to cases where data is "well-behaving" in a sense that it also converges when using CBSEM models instead.

Although the proposed approach is relatively easy to implement with a standard PLS software, can to some extent control for common method variance, and did not produce significant bias when applied to data that were clean of common method variance, more work is needed before this approach can be recommend as an equal alternative to CBSEM based CFA marker variable model for dealing with common method variance. Particularly, the method should be tested with less-optimal Monte Carlo samples and real world data to discover the limits of the approach.

8. References

Agustin, C., & Singh, J. (2005). Curvilinear Effects of Consumer Loyalty Determinants in Relational Exchanges. *Journal of Marketing Research (JMR)*, *42*(1), 96-108.

Alge, B. J., Ballinger, G. A., Tangirala, S., & Oakley, J. L. (2006). Information Privacy in Organizations: Empowering Creative and Extrarole Performance. *Journal of Applied Psychology*, *91*(1), 221-232. doi: 10.1037/0021-9010.91.1.221

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

Burton-Jones, A. (2009). Minimizing method bias through programmatic research. *MIS Quarterly*, *33*(3), 445-A7.

Chin, W. W. (1998a). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295*, 336.

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

Chin, W. W., Marcolin, B. L., & 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), 189-217.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hancock, G. R., & Mueller, R. O. (2006). *Structural equation modeling: a second course*. Information Age Pub Inc.

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2008). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *19*, 1–43.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, *12*(2), 205-218. doi: 10.1037/1082-989X.12.2.205

Kemery, E. R., & Dunlap, W. P. (1986). Partialling Factor Scores Does Not Control Method Variance: A Reply to Podsakoff and Todor. *Journal of Management*, *12*(4), 525.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. Methodology in the social sciences (2nd ed.). New York, NY: The Guilford Press.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114-121.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When" good" indicators are bad and" bad" indicators are good. *Psychological Methods*, *4*, 192–211.

Loehlin, J. (1987). *Latent variable models : an introduction to factor, path, and structural analysis*. Hillsdale N.J.: L. Erlbaum Associates.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology July 2005*, *90*(4), 710-730.

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science*, *52*(12), 1865-1883. doi: 10.1287/mnsc.1060.0597

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

Marcoulides, G. A., & Saunders, C. (2006). PLS: A Silver Bullet? *MIS Quarterly*, *30*(2), 211.

Marsh, H. W., Kit-Tai Hau, Balla, J. R., & Grayson, D. (1998). Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *33*(2), 181.

Maula, M. V. J. (2001). *Corporate venture capital and the value-added for technology-based new firms*. Helsinki University of Technology.

Meehl, P. E., & Waller, N. G. (2002). The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude. *Psychological methods*, *7*(3), 283–300.

Nunnally, J. (1967). *Psychometric theory*. New York: McGraw-Hill.

Pavlou, P. A., & El Sawy, O. A. (2006). From IT Leveraging Competence to Competitive Advantage in Turbulent Environments: The Case of New Product Development. *Information Systems Research*, *17*(3), 198-227.

Pavlou, P. A., Huigang Liang, & Yajiong Xue. (2007). Understanding and mitigating uncertainty in online exchange relationships: a pricipal-agent perspective. *MIS Quarterly*, *31*(1), 105-136.

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

Podsakoff, P. M., MacKenzie, S. B., Jeong-Yeon Lee, & 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), 879.

Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12*(4), 531-544. doi: 10.1177/014920638601200408

Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance- and component-based SEM. *MIS Quarterly*, *33*(1), 197-214.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Lawrence Erlbaum.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Variance. *Organizational Research Methods*, 1094428109332834. doi: 10.1177/1094428109332834

Schepers, J., & Wetzels, M. (2007). A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*, *44*(1), 90–103.

Schwarz, A., Schwarz, C., & Rizzuto, T. (2008). Examining the "Urban Legend" of Common Method Bias: Nine Common Errors and Their Impact. In *Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (p. 441). IEEE Computer Society. Retrieved March 27, 2009, from http://portal.acm.org/citation.cfm?id=1334544

Sharma, R., Yetton, P., & Crawford, J. (2009). Estimating the effect of common method variance: The method-method pair technique with and illustration from TAM research. *MIS Quarterly*, *33*(3), 473-A13.

Shutao Dong, Sean Xin Xu, & Kevin Xiaoguo Zhu. (2009). Information Technology in Supply Chains: The Value of IT-Enabled Resources Under Competition. *Information Systems Research*, *20*(1), 18-32.

Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of AIS*, *2004*(13), 380-427.

Wetzels, M., Odekerken-Schröder, G., & van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, *33*(1), 177-195.

Williams, L. J., & Brown, B. K. (1994). Method Variance in Organizational Behavior and Human Resources Research: Effects on Correlations, Path Coefficients, and Hypothesis Testing. *Organizational Behavior and Human Decision Processes*, *57*(2), 185-209. doi: 10.1006/obhd.1994.1011

Wold, S. (1978). Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics*, *20*(4), 397-405.

Woszczynski, A. B., & Whitman, M. E. (2004). The problem of common method variance in IS research. *The handbook of information systems research*, 66–77.

Ye, J., Marinova, D., & Singh, J. (2007). Strategic Change Implementation and Performance Loss in the Front Lines. *Journal of Marketing*, *71*(4), 156-171.