VARIATION SHARING: A NOVEL NUMERIC SOLUTION TO THE PATH BIAS UNDERESTIMATION PROBLEM OF PLS-BASED SEM

Ned Kock
Shaun Sexton

Full reference:

Kock, N., & Sexton, S. (2017). Variation sharing: A novel numeric solution to the path bias underestimation problem of PLS-based SEM. *International Journal of Strategic Decision Sciences*, 8(4), 46-68.

ABSTRACT

The most fundamental problem currently associated with structural equation modeling employing the partial least squares method is that it does not properly account for measurement error, which often leads to path coefficient estimates that asymptotically converge to values of lower magnitude than the true values. This attenuation phenomenon affects applications in the field of business data analytics; and is in fact a characteristic of composite-based models in general, where latent variables are modeled as exact linear combinations of their indicators. The underestimation is often of around 10% per path in models that meet generally accepted measurement quality assessment criteria. We propose a numeric solution to this problem, which we call the factor-based partial least squares regression (FPLSR) algorithm, whereby variation lost in composites is restored in proportion to measurement error and amount of attenuation. Six variations of the solution are developed based on different reliability measures, and contrasted in Monte Carlo simulations. Our solution is nonparametric and seems to perform generally well with small samples and severely non-normal data.

KEYWORDS: Partial Least Squares; Structural Equation Modeling; Measurement Error; Path Bias; Variation Sharing; Monte Carlo Simulation

INTRODUCTION

Structural equation modeling (SEM) is extensively used in many areas of research, including various business disciplines, as well as the social and behavioral sciences (Kline, 2010; Kock, 2014; Schumacker & Lomax, 2004). The techniques underlying SEM are relevant for the incipient field of business data analytics (Abdelhafez, 2014; Cech et al., 2014; Lee et al., 2014; Liu & Shi, 2015; Wang & Zhou, 2014). SEM employs latent variables, which are measured indirectly through “observed” or “manifest” variables, in sets associated with latent variables that are normally called “indicators”. This measurement includes error. Latent variables typically refer to perception-based constructs (e.g., satisfaction with one’s job). Indicators normally store numeric answers to sets of questions in questionnaires, each set designed to refer to a latent variable, and expected to measure it with a certain degree of imprecision.

Many SEM methods have been proposed over the years. Two main classes of methods have gained wider acceptance: covariance-based and PLS-based SEM (Hair et al., 2011; Kline, 2010; Kock, 2014; Kock & Lynn, 2012). Covariance-based SEM, often viewed as the classic form of SEM, builds on strong parametric assumptions (e.g., multivariate normality) and relies on the minimization of differences between indicator covariance matrices.

PLS-based SEM is generally nonparametric in design, building largely on techniques that make no distributional assumptions. It has a few advantages over covariance-based SEM, such as virtually always converging to solutions; even in complex models, with small sample sizes, and severely non-normal data (Hair et al., 2011; Tenenhaus et al., 2005). Also, PLS-based SEM generates latent variable scores, which can be used in further analyses – e.g., analyses that attempt to uncover and model nonlinear relationships among latent variables (Brewer et al., 2012; Guo et al., 2011; Kock, 2010). Finally, leading software tools for conducting PLS-based SEM (e.g., WarpPLS) tend to be viewed as fairly easy to use by a wide range of researchers.

However, PLS-based SEM builds latent variables as exact linear combinations of their indicators, without explicitly accounting for measurement error. Strictly speaking, these are not really latent variables, but “composites” (McDonald, 1996). Because of this, some argue that PLS-based SEM should not be referred to as an “SEM” technique, while others ignore this as just a semantic issue (Hair et al., 2011). This is one of the reasons why PLS-based SEM is sometimes referred to as “PLS path modeling” (Tenenhaus et al., 2005).

Because PLS-based SEM does not explicitly account for measurement error, it often yields path coefficient estimates that asymptotically converge to values of lower magnitude than the true values as sample sizes grow to infinity. Since path coefficients are proportional to correlations, the amount of underestimation for each path can be approximated through the correlation attenuation factor (Nunnally & Bernstein, 1994), expressed in (1). In this equation, $r(C_i, C_j)$ is the attenuated correlation between composites that refer to two correlated latent variables F_i and F_j; $r(F_i, F_j)$ is the correlation between the latent variables, and α_i and α_j are the true reliabilities associated with the latent variables. We use the symbols F_i and C_i throughout to refer to latent variables (or factors) and associated composites, respectively.

$$r(C_i, C_j) = r(F_i, F_j)\sqrt{\alpha_i \alpha_j}. \quad (1)$$

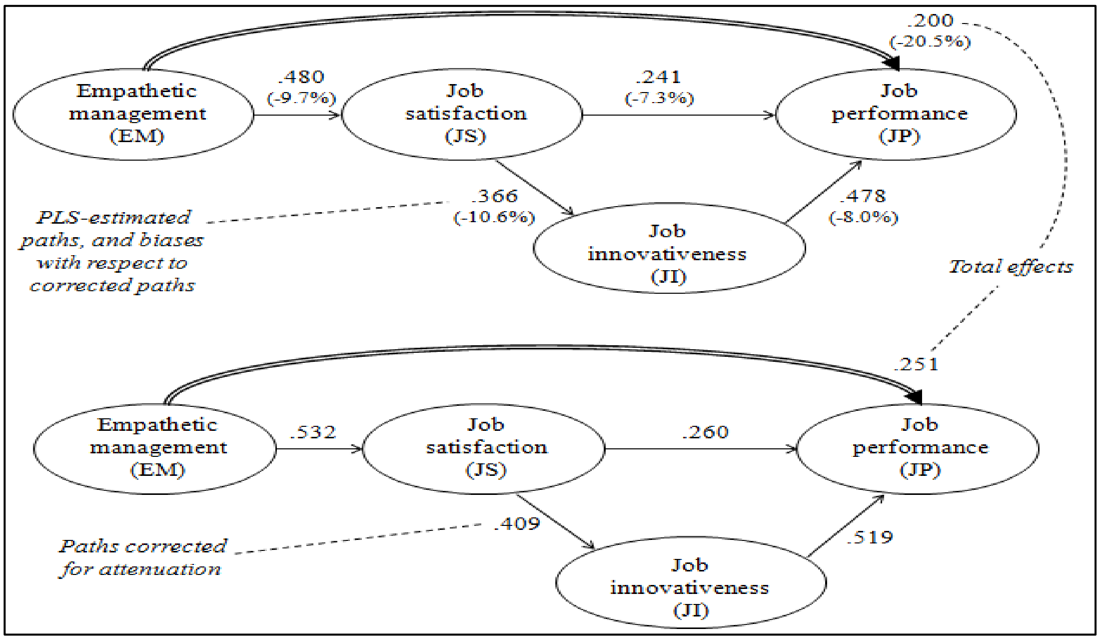
With all its advantages, this path attenuation problem is arguably the "Achilles heel" of PLS-based SEM. Unlike in covariance-based SEM, PLS path estimates tend to converge to absolute values lower than the corresponding absolute true values as sample sizes grow to infinity when a finite number of indicators per latent variable is used. The path coefficients will only converge to the true values if the number of indicators used to measure each factor is also infinite, a property known as "consistency at large" (Cassel et al., 1999).

The top part of Figure 1 shows the results of an analysis that illustrates this path bias problem of PLS-based SEM. It does so in a way that brings to the fore an additional path bias complication, related to total effects. This illustrative analysis is based on a study of the effect of empathetic management (EM) on job performance (JP), which is mediated by intermediate effects on job satisfaction (JS) and job innovativeness (JI).

This study is not presented here as a stand-alone empirical contribution; it is used as an illustration and as the basis for our Monte Carlo simulations, discussed later. The bottom part shows paths corrected based on (1), where Cronbach's alpha coefficients were used as estimates of reliabilities. At the top, path biases are shown as percentages and within parentheses under the PLS-estimated paths. These path biases are based on the corrected paths shown at the bottom.

The data was collected from 257 employees in the southwest region of one of the largest private motor coach charter and schedule service providers in the United States. The latent variables were measured based on question-statements that were previously validated based on other studies, and that also passed validity and reliability criteria in the study in question. This illustrative analysis supports the idea that employee performance is significantly associated with the degree of use of a management style that demonstrates care about the employees' well being (Mayfield & Mayfield, 2009; Mayfield et al, 1998).

Figure 1: Estimated attenuation biases and their impact on direct and total effects



Notes: actual PLS regression path estimates at the top; paths corrected for attenuation bias at the bottom; path biases shown within parentheses at the top, estimated based on corrected paths at the bottom; coefficients at the far right are total effects.

The effect of empathetic management (EM) on job performance (JP) appears to be mediated by intermediate effects on job satisfaction (JS) and job innovativeness (JI), and is thus given by the total effect (coefficients shown at the far right, next to double-lined arrows). The PLS-estimated and corrected total effect coefficients also illustrate a facet of the path bias effect of PLS that makes it more problematic than it looks at first glance; and also more problematic than it has been implied by some researchers primarily based on analyses of direct effects (see, e.g., Cassel et al., 1999). The direct effect path biases can have a cumulative effect on total effects, making the total effect biases much greater than the average direct effect bias. In this illustrative analysis, the total effect bias (-20.5%) is more than twice as large as the average direct effect bias (-8.9%). This gap can be considerably larger in more complex models.

We propose a numeric solution to the path coefficient underestimation problem of PLS-based SEM. Our solution, referred to as the factor-based PLS regression (FPLSR) algorithm, builds on path estimates generated by PLS regression (Wold et al., 2001). It works by essentially restoring measurement error after a PLS-based SEM analysis is completed. Our solution relies on estimates of reliabilities. Given this, we develop six FPLSR variations based on different reliability measures and contrast them in Monte Carlo simulations. The best performer is the one that employs Cronbach's alpha as the reliability measure associated with composites. Our solution is nonparametric and seems to perform well with small samples and severely non-normal data.

Our solution does not rely on parameter correction methods. Examples of solutions that rely on parameter correction methods are the Cronbach alpha disattenuation (Goodhue et al., 2012), illustrated in the example above, and the consistent PLS (Dijkstra & Schermelleh-Engel, 2014) solutions. Generally speaking, parameter correction solutions adjust parameters estimated by classic PLS methods with the goal of obtaining asymptotically unbiased parameters such as path coefficients. Since the FPLSR algorithm yields estimates of factors, and those are used as a basis for estimation of path coefficients, no corrections are needed.

Factor estimation is a characteristic that makes the FPLSR algorithm unique and particularly useful. With factor estimates researchers can employ tests that have been gaining widespread use, and that cannot be conducted without factor estimates. Two notable examples are: (a) full collinearity tests, which concurrently assess both lateral and vertical collinearity among factors (Kock & Lynn, 2012); and (b) nonlinear analyses where best-fitting nonlinear functions are estimated for each pair of linked factors, and subsequently used to estimate path coefficients that take into account the nonlinearity (Guo et al., 2011; Kock, 2010; Moqbel et al., 2013).

PLS-BASED SEM: MAIN ALGORITHMS AND MODES

Four main composite-based estimation algorithms that can be used to analyze models with latent variables are described in this section: PLS mode A (PLSA), PLS mode B (PLSB), PLS regression (PLSR), and standard path analysis (PATH). The latter, PATH, yields composites through a non-iterative procedure; thus it is not, strictly speaking, a PLS algorithm. These algorithms have either been described by or follow directly from Lohmöller (1989), who builds on earlier work by Herman Wold (see, e.g., Wold, 1974). PLSA is the most commonly used in PLS-based SEM.

In all algorithms, indicator weight estimates $\widehat{\omega_{ij}}$ are initially set (usually to 1), and composite estimates $\widehat{C_i}$ are initialized with a standardized vector of the summed indicators. Then the composites are re-estimated based on (2) for PLSA and PLSB; where v_{ij} are referred to as the “inner weights” (Lohmöller, 1989, p. 29), and A_i is the number of composites $\widehat{C_j}$ ($j = 1 \ldots A_i$) that are structurally adjacent to the composite $\widehat{C_i}$. Composites are structurally adjacent when they refer to latent variables that are linked by arrows in the structural model, whether the arrows go in or out. The function $Stdz(\cdot)$ returns a standardized vector. Inner weights are not calculated for PLSR or PATH, as in these two algorithms the indicator weights estimation (PLSR) and direct setting (PATH) do not rely on the estimation of inner weights.

$$\widehat{C_i} = Stdz\left(\sum_{j=1}^{A_i} v_{ij}\, \widehat{C_j}\right). \tag{2}$$

In PLSA and PLSB the inner weights v_{ij} are estimated according to three main schemes: centroid, factorial, and path weighting. The weights v_{ij} are set as the: (a) signs of the correlations among structurally adjacent composites, in the centroid scheme; (b) correlations among structurally adjacent composites, in the factorial scheme; or (c) path coefficients or correlations among structurally adjacent composites, depending on whether the arrows go in or out respectively, in the path weighting scheme.

In the path weighting scheme, which is the most widely used, path coefficient estimates $\widehat{\beta_{ij}}$ are obtained through the solution of (3) for endogenous composites; i.e., composites that have arrows pointing at them in the structural model. Here ζ_i refers to the residual for each endogenous composite, which accounts for the variance that is not explained by the predictor composites $\widehat{C_j}$ ($j = 1 \ldots N_i$) that point at the endogenous composite $\widehat{C_i}$.

$$\widehat{C_i} = \sum_{j=1}^{N_i} \widehat{\beta_{ij}}\, \widehat{C_j} + \zeta_i. \tag{3}$$

Then the following equations are solved for $\widehat{\omega_{ij}}$ (known as "outer weights"), depending on which algorithm is used: (4) is solved in PLSB, and (5) is solved in PLSA and PLSR. In (4) n_i is the numbers of indicators x_{ij} ($j = 1 \ldots n_i$) of the composite $\widehat{C_i}$. Following this, the composites are re-estimated based on (6), a step known as "outside approximation". These and the preceding steps are carried out iteratively until the weight estimates $\widehat{\omega_{ij}}$ change by less than a small fraction.

$$\widehat{C_i} = \sum_{j=1}^{n_i} \widehat{\omega_{ij}}\, x_{ij} + \widehat{\varepsilon_i}. \tag{4}$$

$$x_{ij} = \widehat{\omega_{ij}}\widehat{C_i} + \widehat{\theta_{ij}}. \tag{5}$$

$$\widehat{C_i} \leftarrow Stdz\left(\sum_{j=1}^{n_i} \widehat{\omega_{ij}}\, x_{ij}\right). \tag{6}$$

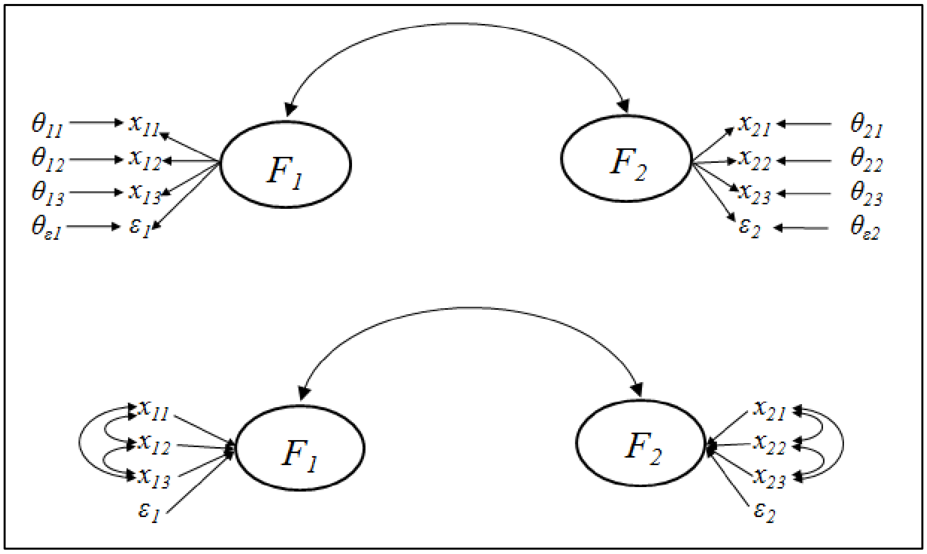
In PATH all of the weights $\widehat{\omega_{ij}}$ are set to 1 and (6) is solved only once. This yields composite estimates that aggregate indicators in an equally-weighted and non-iterative fashion. Given standardization, in PATH the composite estimates are the same for any positive or negative weight assigned to indicators (e.g., .3), as long as all the indicator weights are the same. Setting all weights to 1 is equivalent to estimating composites based on the sums of their indicators.

MEASUREMENT ERROR AND THE ATTENUATION BIAS

Figure 2 shows two correlated factors F_1 and F_2 with three indicators each. Even though the indicators "reflect" their common factors (top part of figure), the factors can also be seen as aggregations of their respective indicators and measurement errors (bottom part of figure). In each factor the measurement error is uncorrelated with the factor's indicators.

Note that even though reflective measurement is assumed, weights do exist and factors can be seen as akin to "composites" that aggregate both indicators and measurement errors. The measurement error that is thus aggregated in each factor could be viewed as an "extra" indicator that: (a) is uncorrelated with the actual indicators; and (b) accounts for the variance in the factor that is not explained by the actual indicators.

Figure 2: Measurement errors for any pair of correlated factors



Note: factors (latent variables) are represented within ovals; the equivalent graph for composites would have the errors ε_1 and ε_2 removed.

This arrangement is expressed for each factor in (7) and (8), where λ_{ij} are the factor-indicator loadings, ω_{ij} are the factor-indicator weights, and n_i is the number of indicators associated with the factor. The indicator measurement errors θ_{ij} arise from the fact that each indicator measures its factor with some degree of imprecision. The measurement error ε_i is due to the errors θ_{ij}, essentially being an aggregated version of those errors. It is the measurement error ε_i that is "left out", or not properly accounted for, when composites are generated in PLS-based SEM.

$$x_{ij} = \lambda_{ij} F_i + \theta_{ij}. \tag{7}$$

$$F_i = \sum_{j=1}^{n_i} \omega_{ij} x_{ij} + \varepsilon_i. \tag{8}$$

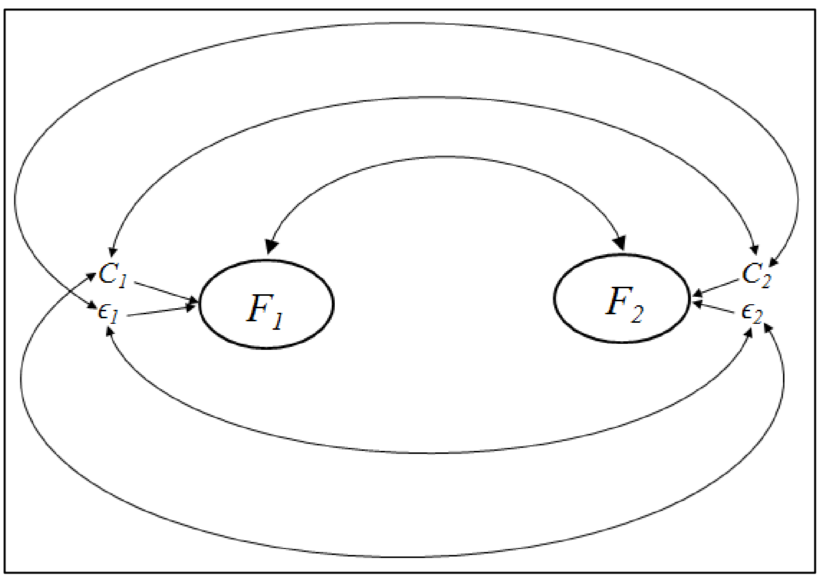
The two correlated factors F_1 and F_2 can be expressed as in (9) and (10); where C_1 and C_2 are the true composites associated with the factors, ω_{1C} and ω_{2C} are the weights of the true composites, ϵ_1 and ϵ_2 are the "base" standardized measurement errors that make up ε_1 and ε_2, and $\omega_{1\epsilon}$ and $\omega_{2\epsilon}$ are the measurement error weights. The true reliabilities α_1 and α_2 equal the corresponding true composite weights squared: ${\omega_{1C}}^2$ and ${\omega_{2C}}^2$ (Nunnally & Bernstein, 1994). Since the standardized base measurement errors and true composites are uncorrelated, it follows that the measurement error weights $\omega_{1\epsilon}$ and $\omega_{2\epsilon}$ equal $\sqrt{1-\alpha_1}$ and $\sqrt{1-\alpha_2}$ respectively.

$$F_1 = \omega_{1C}C_1 + \omega_{1\epsilon}\epsilon_1. \tag{9}$$

$$F_2 = \omega_{2C}C_2 + \omega_{2\epsilon}\epsilon_2. \tag{10}$$

The factors F_1 and F_2 are correlated. Therefore their composites and measurement errors are necessarily cross-correlated, even though composites and measurement errors that refer to the same factor are uncorrelated. That is, even though $r(C_i, \epsilon_i) = 0$; we have $r(C_i, \epsilon_j) \neq 0$, $r(C_i, C_j) \neq 0$ and $r(\epsilon_i, \epsilon_j) \neq 0$. These nonzero cross-correlations are represented in Figure 3.

Figure 3: Cross-correlation of measurement errors and composites



Note: factors, composites, and measurement errors are cross-correlated; but composites and measurement errors that refer to the same factor are not.

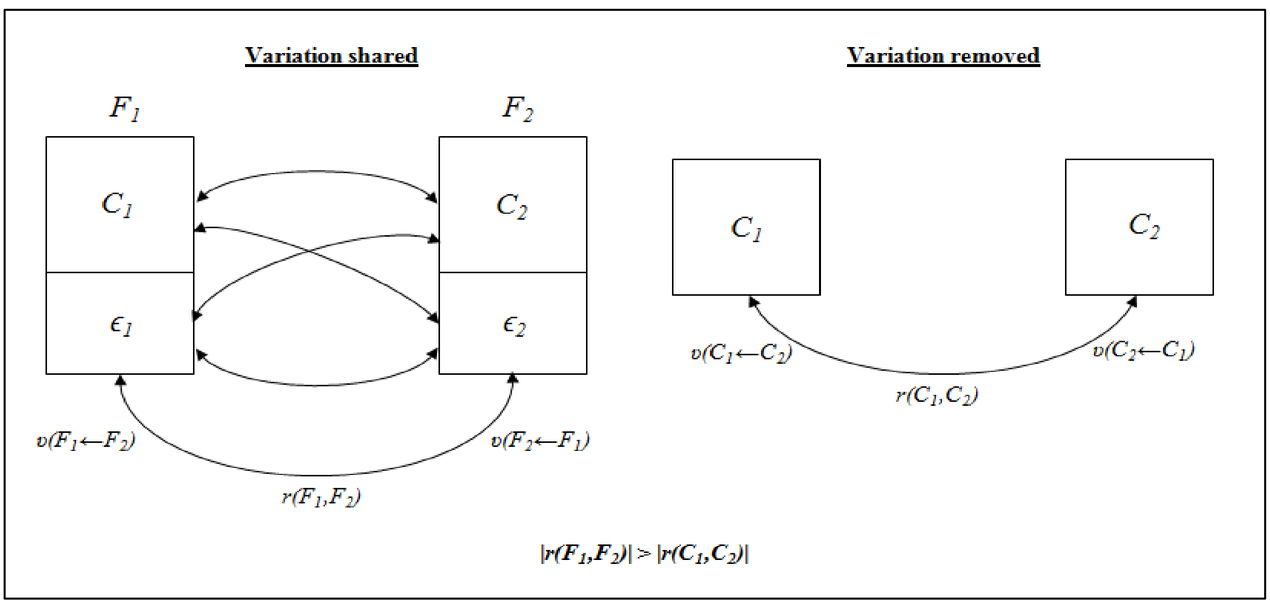
The idea that measurement errors can give rise to an increase in the strength of the correlations between two factors is counterintuitive at first. Generally speaking, the presence of error tends to lead to a decrease in the strength of correlations. The discussion above, however, illustrates why the measurement errors associated with the factors F_1 and F_2 are important in making the strength of the correlation between the factors greater than the strength of the correlation between the corresponding composites.

The nonzero correlations $r(C_1, \epsilon_2)$, $r(C_2, \epsilon_1)$ and $r(\epsilon_1, \epsilon_2)$ contribute additively, together with $r(C_1, C_2)$, to the correlation between the factors $r(F_1, F_2)$. This is why the absolute correlation $|r(C_1, C_2)|$ between the true composites is lower than the absolute correlation $|r(F_1, F_2)|$ between the factors, and ultimately why PLS-based SEM tends to underestimate path coefficients.

VARIATION SHARED AND VARIATION REMOVED

Figure 4 schematically introduces the concepts of "variation shared" and "variation removed", which lead to the property that $|r(F_1, F_2)| > |r(C_1, C_2)|$. There is more variation shared between the factors than between the composites, which causes the correlation between the composites to be attenuated when compared with the correlation between the factors. This variation is shared both ways, from one factor to the other: $v(F_1 \leftarrow F_2)$ and $v(F_2 \leftarrow F_1)$; and from one composite to the other: $v(C_1 \leftarrow C_2)$ and $v(C_2 \leftarrow C_1)$.

Figure 4: Variation shared and variation removed



Note: left – variation shared between factors due to cross-correlations between measurement errors and composites; right – shared variation removed due to removal of measurement errors.

Each factor's measurement error is uncorrelated with the factor's true composite. Therefore, it is clear that the true source of the variation shared between the factors is the shared composite variation. If the true composites associated with any two factors are uncorrelated with each other then the factors will also be uncorrelated. In other words, all "useful" variation emanates from the true composites. Therefore, any loss in shared variation due to composite-based estimation not properly accounting for measurement error can be recovered from the composites.

Let X and Y be two standardized random variables that are correlated with each other. We can always increase the strength of the correlation between X and Y by performing the operation indicated in the left side of (11); where a_1 and a_2 are positive or negative real numbers, depending on whether the initial correlation between X and Y is positive or negative. We refer to this as "sharing variation" between X and Y. The amount of variation shared between X and Y is a function of a_1 and a_2.

$$|r\{Stdz(X + a_1Y), Stdz(Y + a_2X)\}| > |r(X, Y)|. \tag{11}$$

Therefore, we can restore the variation removed from any two correlated composites C_1 and C_2, caused by the removal of correlated measurement errors, by numerically finding appropriate values of a_1 and a_2 that satisfy (12). This would essentially lead to the recovery of the original factors F_1 and F_2 from their true composites.

$$r\{Stdz(C_1 + a_1C_2), Stdz(C_2 + a_2C_1)\} = r(F_1, F_2). \tag{12}$$

From the preceding discussion, it is clear that a_1 and a_2 should be proportional to the measurement error weights $\omega_{1\epsilon}$ and $\omega_{2\epsilon}$ associated with their factors; e.g., a composite derived from a factor that is measured without error should not receive variation from a correlated composite. We can also see that a_1 and a_2 should be proportional to the differences between the factors' and composites' correlations; e.g., if the correlation between composites equals the correlation between factors, no additional variation should be shared between composites. These properties can be expressed more generally, for any pair of composites and factors, through (13) and (14).

$$a_i \propto \omega_{i\epsilon}. \tag{13}$$

$$a_i \propto r(F_i, F_j) - r(C_i, C_j). \tag{14}$$

Equation (15) follows from (13) and (14). In this equation, k is a positive real number that in a numeric fitting algorithm would be expected to minimize the difference between the left and right sides of (12).

$$a_i = k\omega_{i\epsilon}\{r(F_i, F_j) - r(C_i, C_j)\}. \tag{15}$$

There is a different value for the product $\omega_{i\epsilon}\{r(F_i, F_j) - r(C_i, C_j)\}$ for each composite of each pair of correlated composites, and thus a different value of a_i. Given this, k is assumed to stand for a model-wide adjustment quantity that fits a matrix of correlations among composites to the matrix of correlations among factors via a nonparametric equivalent of the expectation-maximization algorithm employed in covariance-based SEM (Moon, 1996; Zhang et al., 2001).

THE PROPOSED FPLSR ALGORITHM AND ITS VARIATIONS

PLSR is equivalent to the commonly used PLSA algorithm; with the exclusion of what is known as PLSA's "inside approximation". In this approximation, weights are assigned to structural paths according to three main schemes – path weighting, centroid, and factorial. This approximation has the effect of creating an interdependence between the inner and outer weights.

Given the absence of the "inside approximation", PLSR yields solutions in which the hypothesized structural model does not influence the estimation of weights. This arguably makes PLSR a more "conservative" model assessment algorithm than PLSA (Kock & Mayfield, 2015), since the researcher does not know the true model prior to the analysis. That is, in PLSR the hypothesized links among factors do not influence the estimation of the composites. For this reason we decided to use PLSR, over other composite-based estimation algorithms, as the basis for our implementation of the proposed FPLSR algorithm.

The steps of the FPLSR algorithm start after PLSR is completed as described earlier. Once stable estimates of composites are obtained via PLSR, the correlation matrices P and S are produced; their elements are indicated as p_{ij} and s_{ij} respectively. The elements of S are obtained via (16). The elements of P are obtained via the application of (17) to the non-diagonal elements of S, which performs an attenuation correction based on (1).

$$s_{ij} = r\left(\widehat{C}_\iota, \widehat{C}_J\right). \tag{16}$$

$$p_{ij} = \frac{s_{ij}}{\sqrt{\widehat{\alpha}_\iota \widehat{\alpha}_J}}. \tag{17}$$

The matrix P stores estimates of the correlations $r\left(F_i, F_j\right)$ among population factors; whereas the matrix S stores the correlations $r\left(\widehat{C}_\iota, \widehat{C}_J\right)$ among the composites initially estimated through the PLSR algorithm. As can be seen, reliability estimates are at the core of our solution. Given this, we conducted preliminary exploratory analyses and subsequently developed the six variations in Table 1 for the estimation of the reliabilities $\widehat{\alpha}_\iota$ and $\widehat{\alpha}_J$. These variations are in turn based on the two most likely candidates for reliability estimates – the Cronbach's alpha and composite reliability coefficients (Aguirre-Urreta et al., 2013; Cronbach, 1951; Peterson & Kim, 2013; Nunnally & Bernstein, 1994; Tenenhaus et al., 2005). In the current study, each of these variations were tested through Monte Carlo simulations. As it will be seen later, the FPLSR-CA variation, building directly on Cronbach's alpha, was the best performer.

In the FPLSR-CA variation, Cronbach's alpha is used directly, without modification, for each factor's estimated reliability. In the FPLSR-CR variation, the composite reliability is used directly. In the FPLSR-AM variation, the arithmetic mean of the Cronbach's alpha and the composite reliability is used. In the FPLSR-SA and FPLSR-SR variations, weighted averages skewed toward the Cronbach's alpha and the composite reliability are used, respectively. In the FPLSR-GM variation, the geometric mean of the Cronbach's alpha and the composite reliability is used.

Once the matrices P and S are generated, each of the composites of a pair of composites $\widehat{C}_\iota$ and $\widehat{C}_J$ that refers to a non-diagonal element of S is iteratively adjusted according to (18). This equation assumes, following our previous discussion, that the adjustment increments are proportional to the weights $\omega_{i\epsilon}$ of the measurement errors ϵ_i (estimated as $\sqrt{1 - \widehat{\alpha}_\iota}$) and the differences between the values of the non-attenuated and attenuated correlations $p_{ij} - s_{ij}$. The elements of the correlation matrix S are re-calculated as indicated in (16) after each adjustment of all of the composites.

$$\widehat{C}_\iota \leftarrow Stdz\left(\widehat{C}_\iota + k\sqrt{1-\widehat{\alpha}_\iota}\left(p_{ij} - s_{ij}\right)\widehat{C}_J\right). \quad (18)$$

Table 1: The six FPLSR variations tested

Variation	Reliability estimate ($\widehat{\alpha}_\iota$) used
FPLSR-CA	$\dot{\alpha}$
FPLSR-CR	$\dot{\rho}$
FPLSR-AM	$\frac{\dot{\alpha}+\dot{\rho}}{2}$
FPLSR-SA	$\frac{2\dot{\alpha}+\dot{\rho}}{3}$
FPLSR-SR	$\frac{\dot{\alpha}+2\dot{\rho}}{3}$
FPLSR-GM	$\sqrt{\dot{\alpha}\dot{\rho}}$

Notes: $\dot{\alpha}$ = Cronbach's alpha for composite; $\dot{\rho}$ = composite reliability.

The value of k is initially set to $1/2$, given that it refers to one of two parts of the variation shared between two composites. As iterations progress, the value of k is reduced to $k/2$, $k/4$ and so on. These reductions are carried out whenever the sum of the absolute differences between the elements of P and S becomes negative. Whenever an adjustment in k is made, the values of S are restored to their most recent values. The iterations continue until the sum of the absolute differences between the elements of P and S changes by less than a small fraction. After these steps are concluded, estimates of the path coefficients $\widehat{\beta_{\iota J}}$ are obtained via (19), where $\widehat{\zeta_\iota}$ refers to the residual for each estimated endogenous factor.

$$\widehat{F_i} = \sum_{j=1}^{N_i} \widehat{\beta_{ij}}\, \widehat{F_j} + \widehat{\zeta_i}. \tag{19}$$

Through this iterative procedure S is fitted to P. This is an entirely numeric and distribution-free solution to the path coefficient underestimation problem of PLS-based SEM. As such, this solution is "true to the PLS tradition", also making no assumptions regarding probability limits of coefficients. It has the effect of restoring the measurement error that is present in (8) but not properly accounted for in (6), by variation sharing among pairs of correlated composites. As such, it generates estimates for factor scores $\widehat{F_i}$ that are no longer composites since they now incorporate measurement error.

MONTE CARLO SIMULATIONS

Monte Carlo simulations (Paxton et al., 2001; Reinartz et al., 2002; Robert & Casella, 2005) were conducted in order to assess our proposed algorithm. Factor scores were generated directly based on a true population model, and indicator scores were subsequently generated based on those factor scores (Goodhue et al., 2012; Mattson, 1997). In these simulations, estimated path coefficients were averaged for sets of 1000 samples (replications). Each set of 1000 samples was generated based on the true population model shown in Figure 5 for the following sample sizes: 50, 100, 200, 300, 400 and 800.

This true population model is based on the actual study mentioned earlier of the effect of empathetic management (EM) on job performance (JP) via intermediate effects on job satisfaction (JS) and job innovativeness (JI). True population path coefficients (.532, .260 etc.) are shown next to arrows, and true population loadings (.900, .850 etc.) are shown next to indicators.

Simulation results generated based on the following algorithms and algorithm variations were contrasted: PATH, PLSA, PLSR, FPLSR-CA, FPLSR-CR, FPLSR-AM, FPLSR-SA, FPLSR-SR, and FPLSR-GM. PLSB was not included because the latent variables are assumed to be reflectively measured, and this algorithm is generally recommended when formative measurement is employed (Lohmöller, 1989).

Figure 5: True population model used in Monte Carlo simulations

Note: the paths in this true model are based on those in the previous illustrative analysis.

Normal and non-normal data were generated. The non-normal data were created based on the power method (Headrick, 2002; 2010) to yield samples with probability limit skewness and excess kurtosis values of 2.828 and 12 respectively. Non-normal latent variable scores and related error terms were created independently from one another to ensure proper non-normality propagation (Kock, 2016). Skewness and excess kurtosis were calculated for all factors and indicators in each of the generated samples to ensure that sample non-normality propagation from factors to indicators occurred properly.

Table 2 shows a summarized set of results, for sample sizes 50 and 300 only. These results include true values of path coefficients, values of path coefficients estimated by the algorithms, and standard deviations of estimates (a.k.a. standard errors). These summarized results are both consistent with and representative of the complete set of results for all of the sample sizes considered. Therefore the complete results are not shown; showing them here would be repetitive. Of the six FPLSR variations, FPLSR-CA was the best performer, and thus only its estimates are shown under the “FPLSR” columns. Both mean path coefficients (rows labeled “Mean”) and standard deviations for path coefficients (rows labeled “SD”) are provided.

These results provide a clear picture of the performance of our proposed FPLSR algorithm with respect to path coefficient bias. The path coefficients generated through this algorithm (variation FPLSR-CA, using Cronbach's alpha) display fairly small biases even with small sample sizes and severely non-normal data.

Table 2: Summarized Monte Carlo simulation results

			Normal data					Non-normal data			
			N=50					N=50			
	True		**PATH**	**PLSA**	**PLSR**	**FPLSR**		**PATH**	**PLSA**	**PLSR**	**FPLSR**
EM→JS (Mean)	.532		.477	.498	.479	.530		.482	.503	.484	.536
EM→JS (SD)			.113	.101	.112	.108		.112	.099	.111	.110
JS→JI (Mean)	.409		.366	.382	.367	.410		.363	.379	.364	.406
JS→JI (SD)			.119	.113	.119	.123		.123	.117	.122	.126
JS→JP (Mean)	.260		.251	.257	.252	.262		.245	.251	.246	.255
JS→JP (SD)			.118	.121	.117	.135		.124	.125	.123	.143
JI→JP (Mean)	.519		.465	.487	.469	.516		.469	.491	.473	.521
JI→JP (SD)			.111	.102	.109	.113		.118	.110	.116	.124
			N=300					**N=300**			
	True		**PATH**	**PLSA**	**PLSR**	**FPLSR**		**PATH**	**PLSA**	**PLSR**	**FPLSR**
EM→JS (Mean)	.532		.479	.485	.481	.532		.480	.485	.481	.532
EM→JS (SD)			.065	.059	.063	.040		.066	.062	.065	.044
JS→JI (Mean)	.409		.367	.372	.368	.411		.365	.370	.367	.409
JS→JI (SD)			.063	.060	.062	.052		.062	.059	.061	.049
JS→JP (Mean)	.260		.251	.253	.252	.261		.248	.251	.249	.257
JS→JP (SD)			.047	.047	.047	.053		.048	.048	.048	.053
JI→JP (Mean)	.519		.471	.479	.475	.522		.471	.479	.475	.522
JI→JP (SD)			.062	.056	.059	.046		.063	.057	.060	.047

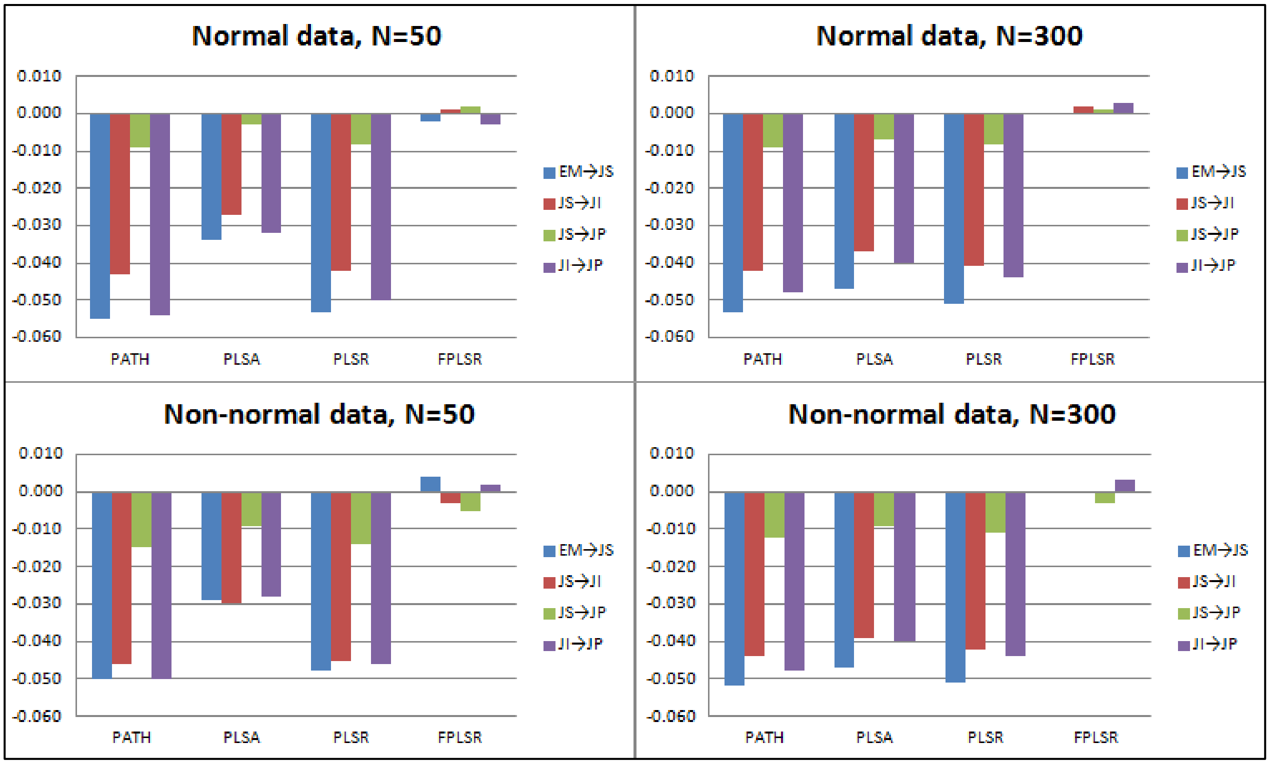
Notes: "Mean" = mean path coefficient; "SD" = standard deviation of path coefficient (a.k.a. standard error); values under the "True" columns are true path coefficients; values under the "FPLSR" columns refer to the FPLSR-CA variation.

At N=50 the standard deviations for the path coefficients increased slightly for FPLSR (with respect to PLSR) in some cases, and decreased slightly in others. At larger sample sizes, these standard deviations consistently decreased. This suggests that the "cost" of the proposed attenuation bias correction is relatively small in terms of variability of path estimates with small samples, and that with larger samples we no longer have a "cost" but a "bonus".

Even in the cases where standard deviations increased, more often than not the increases were offset by increases in the ratios: estimated path coefficient ÷ standard deviation. This is noteworthy because the likelihood of avoiding false negatives is positively correlated with these ratios, and thus so is statistical power. Stated differently, the FPLSR algorithm appears to present the highest power among the algorithms at low sample sizes, whether the data is normally distributed or not.

Figure 6 includes four graphs showing the differences between true and estimated path coefficient values; for sample sizes 50 and 300, as well as normal and non-normal data. The bars at the top-right regions of the graphs, for the FPLSR algorithm, clearly illustrate its remarkable performance in terms of path coefficient estimation precision.

Figure 6: Differences between true and estimated values



Note: bars at top-right region (FPLSR) not visible = perfect matches among estimated and true values.

In those cases where the bars at top-right regions of the graphs in the figure are not visible we have perfect matches among estimated and true path coefficient values. As we can see, the performance of the FPLSR algorithm is far superior to that of the other algorithms even with small samples (N=50) and non-normal data.

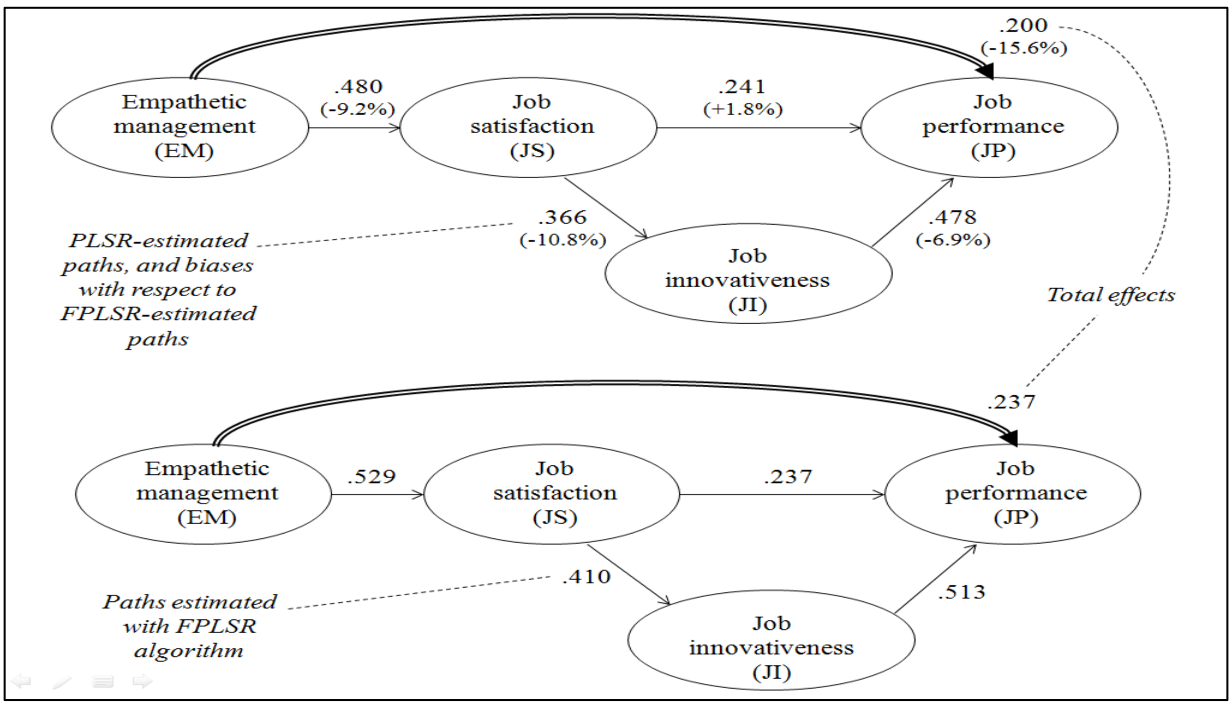
On average the FPLSR algorithm converged after 6 iterations; slightly more than the 5 iterations needed on average for the PLSR algorithm. Combined, the sequence PLSR and FPLSR converged to viable solutions in approximately 11 iterations on average, which can be seen as an indication of good computational efficiency.

EMPIRICAL ILLUSTRATION

We return to our illustrative study mentioned earlier of the effect of empathetic management (EM) on job performance (JP), which is mediated by intermediate effects on job satisfaction (JS) and job innovativeness (JI). In this section we provide a full empirical illustration. As noted before, the data was collected from 257 employees in the southwest region of one of the largest private motor coach charter and schedule service providers in the United States. The question-statements used for data collection are listed in Appendix A.

Figure 7 shows the results of our empirical illustrative analysis, using the PLSR and FPLSR algorithms. For FPLSR, the FPLSR-CA variation is used. Actual PLSR path estimates are shown at the top. The bottom part shows actual FPLSR path estimates. Path biases are shown as percentages and within parentheses under the PLSR-estimated paths. As can be seen, individual path biases have a cumulative effect on the total effect bias, as expected based on our previous discussion of these biases. The average bias associated with direct effects is -6.3%. The total effect bias is over twice as large, at -15.6%.

Figure 7: Results from empirical illustrative analysis



Notes: actual PLSR path estimates at the top; actual FPLSR path estimates at the bottom; FPLSR refers to the FPLSR-CA variation; path biases shown within parentheses at the top; coefficients at the far right are total effects.

We can see that all paths had negative biases associated with them, as expected, with the exception of one – the path from job satisfaction (JS) to job performance (JP). The absolute value of the coefficient estimated for this path was actually higher with PLSR than with our new proposed FPLSR algorithm. For the other paths, the results were fairly compatible with expectations based on our previous discussion.

The anomalous path from job satisfaction (JS) to job performance (JP), which displays a positive bias, illustrates the fact that there is no guarantee that FPLSR-estimated path coefficients will always be of greater magnitude than PLSR-estimated paths, for each and every sample drawn from a population. They will be greater on average, when many samples are considered, but not necessarily in each individual sample among the many samples considered.

Sampling error and specific sample characteristics may cause path coefficients to deviate from common factor model true values (MacCallum & Tucker, 1991). Deviations may include nonzero correlations among structural error terms, which influence correlations among composites, and consequently affect individual path biases. Such nonzero correlations can be seen as indications of hidden confounders that have not been incorporated into the structural model – i.e., they can be seen as indications that the structural model is incomplete.

This seems to be the case in our empirical data, where the correlation between the structural residuals for job satisfaction (JS) and job innovativeness (JI) is -.036. This correlation, which would have been zero in a population model associated with Monte Carlo-generated samples, affects the calculation of path coefficients for the links JS→JI and JS→JP. Since job satisfaction (JS) and job innovativeness (JI) compete for the explained variance in job performance (JP), this nonzero correlation between the structural residuals would tend to decrease the path coefficients for the links JS→JI and JS→JP, compared with the corresponding common factor model true values. This could explain why the FPLSR-estimated values are .237 and .513, and their corresponding common factor model true values are .260 and .519.

It is also interesting to note that job satisfaction (JS) and job performance (JP) appear to be non-normally distributed, which might have compounded the residuals correlation effect. These are suggested by two tests of normality that take as inputs skewness and excess kurtosis values, the classic Jarque-Bera test (Jarque & Bera, 1980; Bera & Jarque, 1981) and Gel & Gastwirth's (2008) robust modification of this test; both of which indicated statistically significantly non-normality for job satisfaction (JS) and job performance (JP). For job satisfaction (JS) skewness is -.521 and excess kurtosis -.591. For job performance (JP) skewness is .134 and excess kurtosis -1.104.

DISCUSSION

In addition to biased path coefficients, loadings are also biased in PLS-based SEM (Reinartz et al., 2009). PLS algorithms tend to yield loadings that are more uniformly distributed than the true population loadings, leading on average to a slight absolute overestimation. In this investigation we focused on the solution of the path coefficient strength underestimation problem of PLS-based SEM. We purposefully avoided the issue of loadings' bias, which we view as a separate issue that is definitely worth tackling but that is nevertheless beyond the scope of this investigation.

Based on exploratory analyses that we have conducted as part of this study and other studies, it appears that the issue of loadings' bias (and, by extension, weights' bias) is the result of nonlinearity among weights and loadings. That is, in data created assuming the common factor model, the relationship between true weights and loadings appears to be nonlinear, with a shape that seems to conform to a logistic function – a type of function that is also found prominently in item response theory models (Hambleton et al., 1991; Waller et al., 2013). PLS algorithms assume by design that weights and loadings are linearly related, and in fact "force" linearity among weights and loadings in the results.

Our solution does not rely on parameter correction methods such as the Cronbach alpha disattenuation (Goodhue et al., 2012) and the consistent PLS (Dijkstra & Schermelleh-Engel, 2014) methods. These methods adjust parameters estimated by classic PLS methods in order to remove bias or make the parameters consistent. Our approach differs from these correction methods in that it yields estimates of factors, from which path coefficients are obtained directly via path analyses. These path coefficients are expected to be estimates of the corresponding true population values. Because of this, in our approach no corrections are needed.

While we contrasted results with normal and severely non-normal data in our analyses, also varying sample size and using a model with path coefficients of different magnitudes, our simulations were limited in scope by the utilization of one single model. More simulations are needed to validate our proposed solution, employing other models. These should include models that are significantly more complex than the one we used, particularly with respect to the network of links among latent variables.

More research is also needed to explore the effects that the variation sharing approach we employed may have on true nonlinear relationships among latent variables. While our approach does not assume that the relationships among composites are linear, it is possible that it may distort underlying nonlinear relationships. The reason for this is that as variation sharing increases the absolute correlations among pairs of estimated composites grows toward unity, a point at which perfect linearity is achieved.

It is not entirely clear whether a significant distortion would occur at the adjustment levels needed to correct for attenuation when psychometrically sound factor indicators are available (i.e., with relatively high loadings). Here variation shared would normally be relatively small. And, if distortion would occur, it is not clear whether it would have a significant impact on nonlinear path coefficient estimation. We conducted preliminary exploratory analyses with the model used in our Monte Carlo simulations including one nonlinear relationship of each of the following types: quadratic, exponential, logarithmic, and logistic. These exploratory analyses suggest that the impact on curve shape and nonlinear path coefficient estimation is likely to be minimal at the adjustment levels needed to correct for attenuation. We intend to pursue this issue further, and recommend it as future research.

CONCLUSION

A variety of SEM methods have been proposed over the years. Two main classes of methods emerged and became more dominant over time in terms of wider academic acceptance. One is covariance-based and the other is PLS-based SEM. Covariance-based SEM, the classic type of SEM, makes strong parametric assumptions (e.g., multivariate normality) and converges to solutions via the minimization of differences between indicator covariance matrices. PLS-based SEM builds largely on techniques that make no distribution assumptions, and virtually always converges to solutions relatively fast; even in complex models, with small sample sizes, and severely non-normal data.

PLS-based SEM implements latent variables as composites, which are exact linear combinations of their indicators. In doing so, it does not explicitly account for measurement error. Because of this PLS-based SEM algorithms generate path coefficient estimates that tend to asymptotically converge to values of lower magnitude than the true values, as sample sizes grow to infinity. This is arguably the “Achilles heel” of PLS-based SEM, which we addressed by proposing a numeric solution, called the factor-based PLS regression (FPLSR) algorithm. Our solution builds on composite estimates generated by PLS regression, which we use as a basis to obtain estimates of the true factors. These estimates explicitly account for measurement error.

Reliability estimates are at the core of our solution. Therefore we developed six FPLSR variations based on various estimates of reliability measures, itself based on the Cronbach's alpha and composite reliability coefficients, and contrasted them in Monte Carlo simulations. The best performer among the variations used Cronbach's alpha directly, without any modification, as the reliability measure associated with composites. Our solution is clearly nonparametric, making no distributional assumptions, and seems to perform generally well with small samples and severely non-normal data.

We are aware that serious questions have been raised regarding Cronbach's alpha's psychometric properties (Sijtsma, 2009), and that our results may be seen as countering those criticisms. This is not exactly the case. It should be noted that our solution uses the Cronbach's alpha coefficients associated with latent variables as a basis for the estimation of what are fundamentally the latent variables' measurement error weights. As such, our solution makes no direct assumptions about the main purported psychometric properties of the Cronbach's alpha coefficient that have been the target of criticism (Sijtsma, 2009). This is an important distinction in light of measurement error theory (Nunnally & Bernstein, 1994), where the Cronbach's alpha coefficient plays a prominent role.

We believe that the nagging problem of path strength underestimation has so far hampered PLS-based SEM broader use among business, social, and behavioral researchers. This has occurred in spite of PLS-based SEM's many advantages, among which is that leading software implementations (e.g., WarpPLS) tend to be viewed as fairly easy to use by a wide variety of researchers. It is our hope that this study will help pave the way for much more extensive use of PLS-based SEM.

We are confident that this study will provide the basis for more targeted research in the future on related topics, which may further establish PLS-based SEM as a viable form of SEM that is complementary to and at the same footing as its classic counterpart – covariance-based SEM. It is our belief that this can be achieved while retaining PLS-based SEM advantages; notably being generally nonparametric in design, and building largely on distribution-free techniques.

REFERENCES

Abdelhafez, H. A. (2014). Big data technologies and analytics: A review of emerging solutions. *International Journal of Business Analytics*, 1(2), 1-17.

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

Bera, A.K., & Jarque, C.M. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, 7(4), 313-318.

Brewer, T.D., Cinner, J.E., Fisher, R., Green, A., & Wilson, S.K. (2012). Market access, population density, and socioeconomic development explain diversity and functional group biomass of coral reef fish assemblages. *Global Environmental Change*, 22(2), 399-406.

Cassel, C., Hackl, P., & Westlund, A.H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, 26(4), 435-446.

Cech, L., Cazier, J., & Roberts, A. B. (2014). Data analytics in the hardwood industry: The impact of automation and optimization on profits, quality, and the environment. *International Journal of Business Analytics*, 1(4), 16-33.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.

Dijkstra, T.K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, 79(4), 585-604.

Gel, Y.R., & Gastwirth, J.L. (2008). A robust modification of the Jarque–Bera test of normality. *Economics Letters*, 99(1), 30-32.

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

Guo, K.H., Yuan, Y., Archer, N.P., & Connelly, C.E. (2011). Understanding nonmalicious security violations in the workplace: A composite behavior model. *Journal of Management Information Systems*, 28(2), 203-236.

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

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Headrick, T. C. (2010). *Statistical simulation: Power method polynomials and other transformations*. Boca Raton, FL: CRC Press.

Headrick, T.C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis*, 40(4), 685-711.

Huang, W. (2013). *PLSe: Efficient estimators and tests for partial least square*. Los Angeles, CA: University of California, Los Angeles.

Jarque, C.M., & Bera, A.K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259.

Kline, R.B. (2010). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.

Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *International Journal of e-Collaboration*, 6(4), 1-11.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1-13.
Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, 15(1), 299-315.
Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.
Kock, N., & Mayfield, M. (2015). PLS-based SEM algorithms: The good neighbor assumption, collinearity, and nonlinearity. *Information Management and Business Review*, 7(2), 113-130.
Lee, Y. M., An, L., Liu, F., Horesh, R., Chae, Y. T., & Zhang, R. (2014). Analytics for smarter buildings. *International Journal of Business Analytics*, 1(1), 1-15.
Liu, K., & Shi, J. (2015). A systematic approach for business data analytics with a real case study. *International Journal of Business Analytics*, 2(4), 23-44.
Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg, Germany: Physica-Verlag.
MacCallum, R.C., & Tucker, L.R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, 109(3), 502-511.
Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, 32(4), 355-373.
Mayfield, J., & Mayfield, M. (2009). The role of leader motivating language in employee absenteeism. *International Journal of Business Communication*, 46(4), 455–479.
Mayfield, J., Mayfield, M., & Kopf, J. (1998). The effects of leader motivating language on subordinate performance and satisfaction. *Human Resource Management*, 37(3-4), 235–248.
McDonald, R.P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, 31(2), 239-270.
Moon, T.K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6), 47-60.
Moqbel, M., Nevo, S., & Kock, N. (2013). Organizational members' use of social networking sites and job performance: An exploratory study. *Information Technology & People*, 26(3), 240-264.
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
Paxton, P., Curran, P.J., Bollen, K.A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2), 287-312.
Peterson, R.A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98(1), 194-198.
Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332-344.
Reinartz, W.J., Echambadi, R., & Chin, W.W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research*, 37(2), 227-244.
Robert, C.P., & Casella, G. (2005). *Monte Carlo statistical methods*. New York, NY: Springer.
Schumacker, R.E., & Lomax, R.G. (2004). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.

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

Waller, J., Ostini, R., Marlow, L.A., McCaffery, K., & Zimet, G. (2013). Validation of a measure of knowledge about human papillomavirus (HPV) using item response theory and classical test theory. *Preventive Medicine*, 56(1), 35-40.

Wang, J. & Zhou, S. B. (2014). Making data-driven discerning decision with business analytics. *International Journal of Business Analytics*, 1(1), iv-vii.

Wold, H. (1974). Causal flows with latent variables: Partings of the ways in the light of NIPALS modelling. *European Economic Review*, 5(1), 67-86.

Wold, S., Trygg, J., Berglund, A., & Antti, H. (2001). Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 131-150.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45-57.

APPENDIX A: QUESTIONS USED IN ILLUSTRATIVE ANALYSIS

The question-statements below were used for latent variable measurement in the study depicted in the illustrate analysis. Question-statements were answered on 5-point Likert-type scales.

Empathetic management (EM)

- EM1: My supervisor gives me praise for my good work.
- EM2: My supervisor shows me encouragement for my work efforts.
- EM3: My supervisor shows concern about my job satisfaction.
- EM4: My supervisor expresses his/her support for my professional development.
- EM5: My supervisor shows trust in me.

Job satisfaction (JS)

- JS1: I always feel satisfied with my job.
- JS2: I like my job.
- JS3: I do not want to change my job.
- JS4: I like my job more than others.
- JS5: I like telling people about my job.

Job innovativeness (JI)

- JI1: I try new ideas and approaches to problems.
- JI2: I welcome uncertainty and unusual circumstances related to my tasks.
- JI3: I can be counted on to find a new use for existing methods or equipment.
- JI4: I demonstrate originality.
- JI5: I provide critical input toward a new solution.

Job performance (JP)

- JP1: Which of the following selections best describes how your supervisor rated you on your last formal performance evaluation?
- JP2: How does your level of production quantity compare to that of your colleagues' productivity levels?
- JP3: How does the quality of your products or services compare to your colleagues' output?
- JP4: How efficiently do you work compared to your colleagues? In other words, how well do you use available resources (money, people, equipment, etc.)?
- JP5: Compared to your colleagues, how good are you at preventing or minimizing potential work problems before they occur?
- JP6: Compared to your colleagues, how effective are you with keeping up with changes that could affect the way you work?
- JP7: How quickly do you adjust to work changes compared to your colleagues?
- JP8: How well would you rate yourself compared to your colleagues in adjusting to new work changes?
- JP9: How well do you handle work place emergencies (such as crisis deadlines, unexpected personnel issues, resources allocation problems, etc.) compared to your colleagues?