Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods

Ned Kock
Pierre Hadaya

Abstract

Partial least squares-based structural equation modeling (PLS-SEM) is extensively used in the field of information systems, as well as in many other fields where multivariate statistical methods are employed. One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation. The "10-times rule" has been a favorite due to its simplicity of application, even though it tends to yield imprecise estimates. We propose two related methods, based on mathematical equations, as alternatives for minimum sample size estimation in PLS-SEM: the inverse square root method, and the gamma-exponential method. Based on three Monte Carlo experiments, we demonstrate that both methods are fairly accurate. The inverse square root method is particularly attractive in terms of its simplicity of application.

Keywords: Information Systems, Partial Least Squares, Structural Equation Modeling, Statistical Power, Monte Carlo Simulation

Introduction

The field of information systems (IS) is closely linked with the development, software implementation, and use of the partial least squares (PLS) technique (Chin, 1998; Chin et al., 2003; Kock, 2010; Wold, 1980). This technique has been extensively used in IS (Bradley et al., 2012; Goodhue et al., 2012), a practice that has extended to other fields over the years, to analyze path models with variables that are measured indirectly through other variables (Chin, 1998; Hair et al., 2011; 2014; Lohmöller, 1989). These indirectly measured variables are generally known as latent variables (Kline, 1998; Kock & Lynn, 2012). The approach to analyzing path models with latent variables is broadly known as structural equation modeling (SEM). Thus the acronym "PLS-SEM" is used here to refer to SEM employing PLS.

One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation. A widely used minimum sample size estimation method in PLS-SEM is the "10-times rule" method (Hair et al., 2011), which builds on the assumption that the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model. While this method's simplicity of application makes it a favorite among PLS-SEM users, it has been shown in the past to lead to inaccurate estimates (Goodhue et al., 2012).

We propose two related methods, based on mathematical equations, for minimum sample size estimation in PLS-SEM. The first method is called the inverse square root method, because it uses the inverse square root of a sample's size for standard error estimation – an important step in minimum sample size estimation. The second method is called the gamma-exponential method, since it relies on gamma and exponential smoothing function corrections applied to the first method. Based on Monte Carlo experiments, we show that both methods are fairly accurate, with the first method being also particularly attractive in terms of its simplicity of application.

The methods we propose here should be seen as heuristic methods; that is, as practical methods that are not guaranteed to yield optimal estimates. We believe that they are valuable time-saving tools to be used in the early cyclical phases of research design, and that can be significantly more precise than comparable early-stage research heuristics currently used by researchers. More specifically, they are first-step methods that researchers can employ to address issues related to statistical power and minimum sample size requirements (Cohen, 1988; 1992; Goodhue et al., 2012; Kock, 2016).

We use a unique study in the field of IS to illustrate our discussion of minimum sample size estimation in PLS-SEM. The study targeted was authored by Majchrzak, Beath, Lim, and Chin (MBLC), and published in the journal *MIS Quarterly* in 2005 (Majchrzak et al., 2005). MBLC's study, which investigated a collaborative approach for IS design, apparently has the distinction of being the PLS-SEM study published in an elite IS research journal with the smallest sample size ever. It employed a sample size of 17.

MBLC's study

MBLC's study focused on the impact that a cooperative learning strategy called collaborative elaboration, developed by educational psychologists, had on client learning and ultimately on short- and long-term outcomes in the context of IS design projects. Figure 1 shows the model that they used, with the main results of their analyses. The path estimation algorithm they employed was PLS Mode A (Lohmöller, 1989), and the P value calculation method was

bootstrapping (Diaconis & Efron, 1983; Efron et al., 2004). This algorithm and P value calculation method are by far the most widely used in PLS-SEM.

Figure 1: The model in MBLC's study with results
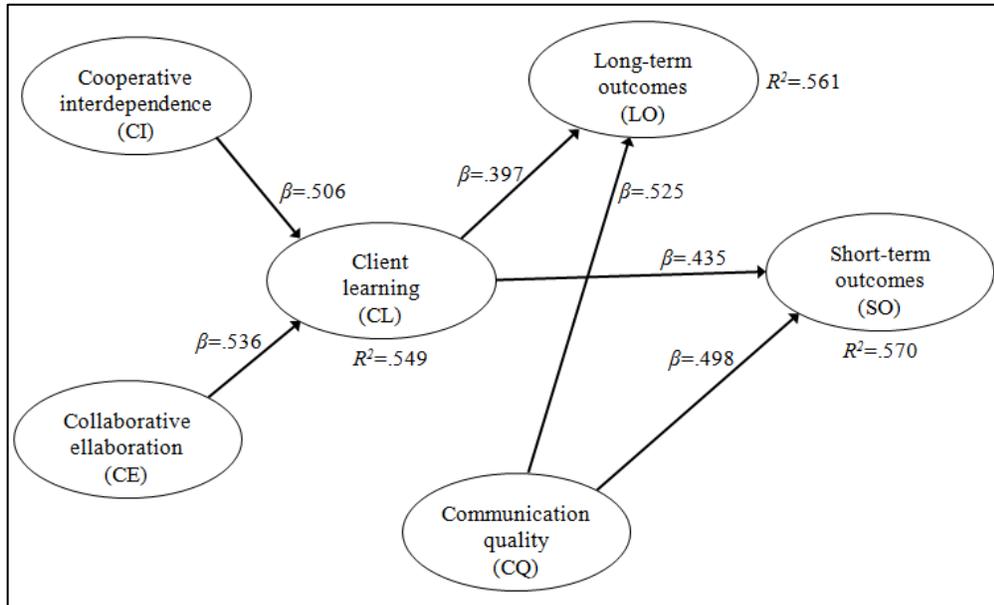
The latent variables shown as ovals were measured reflectively through multiple indicators, primarily on Likert-type scales with 5 points. The question-statements associated with each indicator were answered with respect to project meetings. Cooperative interdependence (CI) was measured based on 2 indicators and assessed the degree to which the tasks planned for the meeting were in fact accomplished. Collaborative elaboration (CE) was measured based on 6 indicators and assessed the degree to which clients and developers elaborated on their respective views about their projects. Client learning (CL) was measured based on 3 indicators and assessed the degree to which clients changed their views about the IS project requirements. Developers' communication quality (CQ) was measured based on 7 indicators and assessed the degree to which developers were good communicators. Long-term outcomes (LO) was measured based on 4 indicators and assessed the degree to which clients expanded their knowledge about IS and their development. Short-term outcomes (SO) was measured based on 4 indicators and assessed the degree to which the project would lead to a successful IS implementation.

MBLC collected data from 17 project teams comprising 68 developers and 17 clients. Each team had 4 developers and 1 client. The 17 teams met three times during the project, which lasted 12 weeks, and which culminated with the development of a IS prototype. Immediately following each of the three meetings, the clients were surveyed about the degree of CI exhibited during the meeting, the use of CE by the developers and themselves, as well as the extent to which CL occurred. The assessments of CI, CE and CL were then averaged across the 3 meetings in order to obtain richer and more stable measures. In addition to the meeting assessments, the clients were also surveyed on the outcomes of the IS design phase (LO and SO) and CQ at the conclusion of the 12-week project.

MBLC's study was based on a solid theoretical development, and, as mentioned before, was published in the journal *MIS Quarterly*, which has long been considered a very selective elite

academic IS journal. We use their study as a basis to contrast our proposed methods for minimum required sample size estimation against existing comparison methods.

Power, effect size, and minimum sample size

Statistical **power** (Cohen, 1988; 1992; Goodhue et al., 2012; Kock, 2016; Muthén & Muthén, 2002), often referred to simply as "power", is a statistical test's probability of avoiding type II errors, or false negatives. Power is often estimated for a particular coefficient of association and sample size, for samples drawn from a population, at a given significance level (usually P < .05). For example, let us consider a PLS-SEM test employing PLS Mode A and bootstrapping. Let us assume that such a test is able to recognize a path coefficient as statistically significant, where the path coefficient is associated with a "real" effect at the population level of magnitude .2; which would be referred to as the "true" path coefficient. Let us also assume that the test correctly recognizes the path coefficient as significant 83 percent of the time when samples of size 150 are randomly taken from the population. Under these circumstances, we would conclude that the power of the test is 83 percent, or .83.

The **effect size** (Cohen, 1988; 1992; Kock, 2014b) is a measure of the magnitude of an effect that is independent of the size of the sample analyzed. Two main measures of effect size are commonly used in PLS-SEM. The most widely used is Cohen's f^2 coefficient (Cohen, 1988; 1992), which is calculated as $\Delta R^2 / (1 - R^2)$, where ΔR^2 is the incremental contribution of a predictor latent variable to the R^2 of the criterion latent variable to which it points. The other measure of effect size commonly used in PLS-SEM is the absolute contribution of the predictor latent variable (Kock, 2014b; Mandal et al., 2012), namely the numerator ΔR^2 of Cohen's f^2 equation, without the denominator correction. This second measure tends to yield lower results, thus being a more conservative effect size estimate. By convention, effect sizes of 0.02, 0.15, and 0.35 are respectively termed small, medium, and large (Cohen, 1992; Kock, 2014b).

The **minimum sample size** at which a PLS-SEM test achieves an acceptable level of power (usually .8) depends on the effect size associated with the path coefficient under consideration (Cohen, 1988; 1992; Goodhue et al., 2012; Kock, 2014b). The higher is the magnitude of a path coefficient at the population level, the higher is usually its effect size, and the greater is the probability that a true effect will be properly detected with a small sample. Therefore strong path coefficients at the population level, whether they are negative or positive, tend to require very small sample sizes for their proper identification. So, if a researcher knows that all of the path coefficients of a model will be strong prior to collecting empirical data, leading to large effect sizes, the researcher may consider using a small sample size in a PLS-SEM analysis. As we will see later, we can use the notion of effect size for a general minimum sample size recommendation that does not rely on predictions about path strength.

More often than not PLS-SEM is presented as being a desirable multivariate data analysis method due to its remarkable ability to achieve acceptable power at very small sample sizes (Hair et al., 2011; 2014). While this may be true for models containing only strong path coefficients and large effect sizes, it is not true for models with path coefficients of more modest magnitudes, and certainly not true for models with fairly weak path coefficients. (At points in our discussion we deviate somewhat from strict technical statistical jargon, for simplicity. For example, in the previous sentence we refer to "weak" path coefficients, meaning positive or negative path coefficients whose absolute values are low.) It will be demonstrated here that PLS-

SEM's power is consistent with what one would expect from ordinary least squares regression, and probably other methods with similar mathematical underpinnings.

Comparison methods for minimum sample size estimation

In this section we discuss three methods for minimum sample size estimation in PLS-SEM that we use as a basis for comparison when we evaluate our proposed methods. The first method presented here relies on Monte Carlo simulations (Paxton et al., 2001; Robert & Casella, 2013). The second method, the 10-times rule method (Goodhue et al., 2012; Hair et al., 2011), is the most widely used in PLS-SEM, in the field of IS as well as other fields. The third method, the minimum R-squared method, has been proposed by Hair et al. (2014, p. 21) as an alternative to the 10-times rule method.

The Monte Carlo simulation method

Employing the Monte Carlo simulation (Kock, 2016; Paxton et al., 2001; Robert & Casella, 2013) method for minimum sample size estimation in PLS-SEM requires the researcher to set a number of sample size points (e.g., 15, 20, 30 and 40), generate a number of samples (e.g., 1000) for each sample size point, calculate the percentages of samples in which significant effects (e.g., for which $P < .05$) were found for each sample size point (the power associated with each sample size), and estimate via interpolation the minimum sample size at which power reaches the desired threshold (i.e., .8). Table 1 illustrates this process through a set of results for four sample size points.

The table shows the power values calculated for each sample size for the CL→LO link in the model used in MBLC's study, whose path coefficient was estimated at .397. The process has to be performed individually for each path coefficient in a PLS-SEM model. From the table, we can see that the power threshold of .8 is reached for a sample size N where $20 < N < 30$. Through simple linear interpolation we calculate the smallest positive integer greater than

$$20 + (30 - 20)\frac{(.8-.646)}{(.847-.646)}, \text{ which is } 28.$$

Thus the minimum required sample size is estimated via this method to be 28. Note that this method relies on a well informed choice of sample size points, which in this example are: 15, 20, 30 and 40. Another set of sample size points may not encompass the sample size for which the power threshold of .8 is reached; for example: 100, 200, 300 and 400.

Table 1: The Monte Carlo simulation method

N	Power
15	.477
20	.646
30	.847
40	.919

Therefore, the Monte Carlo simulation method often requires two or more simulations. The initial simulations are experimental, to define an appropriate set of sample size points. These

would be followed by a final simulation, whereby one would estimate via interpolation the minimum sample size at which power reaches the desired threshold of .8.

The samples (e.g., 1000) generated for each sample size point via the Monte Carlo simulation method are based on a population model defined by the researcher. The process of building samples in the Monte Carlo simulation method also relies on common factor model assumptions (Kline, 1998; Kock, 2016). This process is explained in more detail in Appendix A. The Monte Carlo simulation method is a complex way by which minimum sample sizes can be determined, and for which technical methodological expertise is required.

As we can see, minimum sample size estimation via the Monte Carlo simulation method may be a very time consuming alternative, even for experienced methodological researchers with good computer programming skills. Nevertheless, it is a fairly precise method for minimum sample size estimation, and in fact the preferred method for that purpose. As such, we use it to obtain baseline estimates against which other methods are compared.

The 10-times rule method

The most widely used minimum sample size estimation method in PLS-SEM, in the field of IS as well as other fields, is the "10-times rule" method (Hair et al., 2011; Peng & Lai, 2012). Among the variations of this method, the most commonly seen is based on the rule that the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model (Goodhue et al., 2012).

Unlike in the Monte Carlo simulation method, minimum sample size estimation via the 10-times rule method does not depend on the magnitude of the path coefficients in the model. For example, in the model used in MBLC's study, the 10-times rule method leads to the minimum sample size estimation of 20, regardless of the strengths of the path coefficients. This is because the maximum number of model links pointing at any variable in the model is 2, which multiplied by 10 yields 20. As we will see later, this method can lead to grossly inaccurate estimations of minimum required sample size.

The minimum R-squared method

In their pioneering book on PLS-SEM, Hair et al. (2014, p. 21) discuss an alternative to the 10-times rule for minimum sample size estimation. We refer to this method as the "minimum R-squared method", because the minimum R^2 in the model is prominently used for minimum sample size estimation. This method, which builds on Cohen's (1988; 1992) power tables for least squares regression, relies on a table listing minimum required sample sizes based on three elements.

The first element of the minimum R-squared method is the maximum number of arrows pointing at a latent variable (a.k.a. construct) in a model. The second is the significance level used. The third is the minimum R^2 in the model. Table 2 is a reduced version of the table presented by Hair et al. (2014, p. 21). This reduced version focuses on the significance level of .05, which is the most commonly used significance level in the field of IS, and assumes that power is set at .8.

For example, in the model used in MBLC's study, the maximum number of arrows pointing at a latent variable is 2, and the minimum R^2 in the model is .549. There is no cell in the table for the minimum R-squared method for which these two values intersect, but the closest cell shows a minimum sample size of 33, which we use as the estimate. As we can see, this method appears to be an improvement over the 10-times rule method, as it takes as an input at least one additional

element beyond the network of links in the model. However, this method (i.e., the minimum R-squared method) can also lead to grossly inaccurate estimations of minimum required sample size, which we will see later.

Table 2: Table for the minimum R-squared method

Maximum number of arrows pointing at a construct	Minimum R^2 in the model			
	.10	**.25**	**.50**	**.75**
2	110	52	33	26
3	124	59	38	30
4	137	65	42	33
5	147	70	45	36
6	157	75	48	39
7	166	80	51	41
8	174	84	54	44
9	181	88	57	46
10	189	91	59	48

Our proposed methods for minimum sample size estimation

In this section we discuss two related methods, based on mathematical equations, for minimum sample size estimation in PLS-SEM. Neither method relies on Monte Carlo simulations or on elements that make up the 10 times rule or the minimum R-squared methods. The first method, called the inverse square root method, uses the inverse square root of a sample's size for standard error estimation – hence its name. The second method, called the gamma-exponential method, relies on gamma and exponential smoothing function corrections applied to the standard error estimation employed in the first method.

The inverse square root method

Whenever one or more researchers analyze samples taken from a population using PLS-SEM, each analysis generates various path coefficients. Each path coefficient (β) will have a standard error (S) associated with it. If we plot the distribution of the ratio β/S, also indicating the location of a critical T ratio (Kock, 2015; Weakliem, 2016) for a specific significance level chosen, we will obtain a graph that will have the general shape shown in Figure 2. For each instance where the ratio β/S surpasses the critical T ratio, the effect associated with the path coefficient β will be correctly deemed as statistically significant. This assumes that the path coefficient refers to an effect that exists at the population level – a "true" effect.

The magnitude of the ratio β/S increases with increases in the magnitude of the path coefficient β and decreases in the standard error S. This standard error decreases with increases in sample size, as will be seen shortly below. Therefore, with increases in the magnitude of the path coefficient and of the sample size analyzed, the probability that the ratio β/S will surpass the critical T ratio will increase. As a result, the likelihood that an effect that does exist at the population level will be mistakenly rejected will decrease. In other words, the power of the test will increase.

Figure 2: Distribution of the ratio β/S
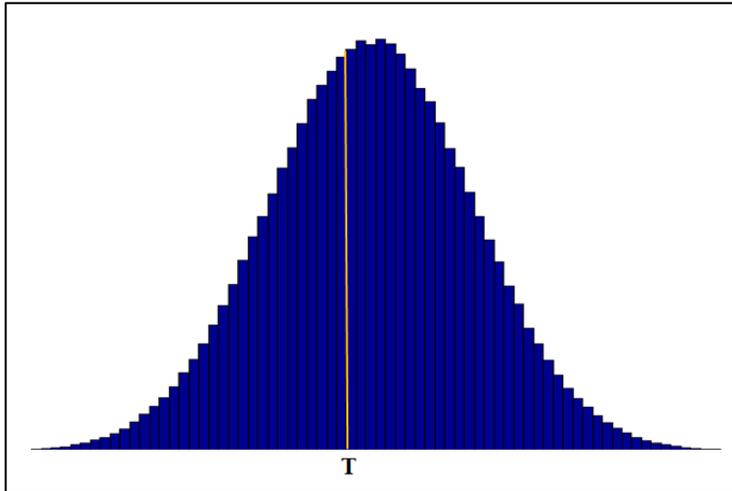
Note: T = critical T ratio for a specific significance level chosen.

As we can see from the figure, the power of a test associated with a given path coefficient for which a sign has been hypothesized can be defined as the probability that the ratio $|\beta|/S$ will be greater than the critical T ratio for a specific significance level chosen (Cohen, 1988; Goodhue et al., 2012; Kock, 2015). Here $|\beta|$ is the absolute value of β, as a path coefficient strength's influence on power is exerted whether the coefficient is positive or negative. The significance level normally chosen in the field of IS is .05 (i.e., P < .05), for which the critical T ratio can be denoted as $T_{.05}$. This can be expressed mathematically as follows.

$$W = P\left(\frac{|\beta|}{S} > T_{.05}\right). \tag{1}$$

Statistical power is denoted as W in (1), and $P(\cdot)$ is the probability function. If we set power to be above a given level, most commonly .8 in IS research, the above can be expressed employing a cumulative probability function $\Phi(\cdot)$ for the standard normal distribution. Assuming that path coefficients are normally distributed, we can say that power will be greater than .8 when the cumulative distribution function for the standard normal distribution indicated in (2) is greater than .8.

$$\Phi\left(\frac{|\beta|}{S} - T_{.05}\right) > .8. \tag{2}$$

The assumption that path coefficients are normally distributed generally holds for PLS-SEM, because coefficients calculated based on sample sets taken randomly from a population tend to be distributed in conformity with the central limit theorem (Kipnis & Varadhan, 1986; Miller & Wichern, 1977).

Taking (2) as a basis, we obtain (3) in terms of the standardized score associated with the value .8 of the cumulative distribution function for the normal distribution ($z_{.8}$). To obtain (3) we also take into consideration the property that $T_{.05} = z_{.95}$.

$$\frac{|\beta|}{S} - T_{.05} > z_{.8} \rightarrow$$

$$\frac{|\beta|}{S} > T_{.05} + z_{.8} \rightarrow$$

$$\frac{|\beta|}{S} > z_{.95} + z_{.8}. \tag{3}$$

Any given z-score z_x can be calculated based on a standard normal distribution, which is a normal distribution with a mean of 0 and a standard deviation of 1. The score is a value associated with the probability x that a random variable takes on a value that is equal to or less than z_x. In MATLAB it is obtained using the function norminv(x,0,1). In Excel it is obtained using the function NORMINV(x,0,1) or the function NORMSINV(x).

An estimate $\hat{S}$ of the true standard error (S) can be produced through (4). This estimate lends the name to the method presented here, the inverse square root method, and is known to be biased (Gurland & Tripathi, 1971; Kock, 2014a), consistently underestimating the corresponding true value at very small samples (i.e., $1 < N \leq 10$), and consistently overestimating it at greater sample sizes (i.e., $N > 10$). Shortly we will discuss two approaches to correct this bias, which are combined in our second proposed minimum sample size estimation method, the gamma-exponential method.

$$\hat{S} = \frac{1}{\sqrt{N}}. \tag{4}$$

Employing the Excel function NORMSINV(x) we obtain the values for $z_{.95}$ and $z_{.8}$, or NORMSINV(.95) and NORMSINV(.8), which are respectively 1.645 and 0.842. The sum $z_{.95} + z_{.8}$ is thus 2.486. Combining (3) and (4), with $|\beta|_{min}$ replacing $|\beta|$ and representing the absolute value of the statistically significant path coefficient with the minimum magnitude in the model, we then have:

$$|\beta|_{min}\sqrt{\hat{N}} > z_{.95} + z_{.8} \rightarrow$$

$$\hat{N} > \left(\frac{z_{.95} + z_{.8}}{|\beta|_{min}}\right)^2 \rightarrow$$

$$\hat{N} > \left(\frac{2.486}{|\beta|_{min}}\right)^2. \tag{5}$$

Based on our proposed inverse square root method, the minimum sample size is estimated as the smallest positive integer that satisfies (5). As such, it can be calculated by rounding the result of the calculation of the right side of the equation to the next integer. In MATLAB it can be obtained using the function ceil((2.486/bmin)^2), where bmin is a variable that stores the value of $|\beta|_{min}$. In Excel it can be obtained using the function ROUNDUP((2.486/bmin)^2,0), where bmin is the name of a cell that stores the value of $|\beta|_{min}$.

The gamma-exponential method

As we noted earlier, our estimate $\hat{S}$ of the true standard error (S), obtained through the formula $1/\sqrt{N}$, is known to be biased. A classic gamma function correction of the bias for very small sample sizes (i.e., $1 < N \leq 10$) was proposed by Gurland & Tripathi (1971):

$$\hat{S} = \frac{1}{c\sqrt{N}},$$

where

$$c = \sqrt{\frac{N-1}{2}} \; \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}, \text{ and } \Gamma(\cdot) \text{ is the gamma function.}$$

With the gamma function correction proposed by Gurland & Tripathi's (1971), the resulting equation (6) to obtain the minimum required sample size $\widehat{N}$ becomes more complex. This equation can be solved by means of a computer program that starts with $\widehat{N} = 1$ and progressive increments the value of $\widehat{N}$ to 2, 3 etc. until the smallest positive integer that satisfies the equation is obtained. In MATLAB the value of $\Gamma(x)$ is obtained using the function gamma(x). In Excel it is obtained using the two-function formula EXP(GAMMALN(x+1)).

$$|\beta|_{min}\sqrt{\widehat{N}}\sqrt{\frac{\widehat{N}-1}{2}} \; \frac{\Gamma\left(\frac{\widehat{N}-1}{2}\right)}{\Gamma\left(\frac{\widehat{N}}{2}\right)} > 2.486. \tag{6}$$

The gamma function correction equation has no effect, in terms of minimum required sample size estimation, for $N > 10$. The reason for this is that the correction coefficient c quickly converges to 1 for $N > 10$. An exponential smoothing function correction of the standard error bias was proposed and validated by Kock (2014a) in the context of PLS-SEM for sample sizes greater than those covered by the gamma function correction (i.e., $N > 10$):

$$\hat{S} = \frac{1}{\sqrt{N}} e^{-\left(\frac{e|\beta|}{\sqrt{N}}\right)}.$$

With this exponential smoothing function correction, the equation (7) to obtain the minimum required sample size $\widehat{N}$ also ends up being more complex. As with the gamma function correction equation, this equation can be solved with a computer program that starts with $\widehat{N} = 1$ and progressive increments its value to 2, 3 etc. until the smallest positive integer that satisfies the equation is obtained. In MATLAB the value of e^x is obtained using the function exp(x). In Excel it is obtained using the function EXP(x).

$$|\beta|_{min}\sqrt{\widehat{N}}e^{\left(\frac{e|\beta|_{min}}{\sqrt{\widehat{N}}}\right)} > 2.486. \tag{7}$$

We developed an Excel spreadsheet with Visual Basic code, discussed in Appendix B, to obtain corrected estimates based on equations (6) and (7). This enabled us to implement the gamma-exponential method, by combining gamma and exponential smoothing function corrections applied to the standard error estimation employed in the inverse square root method. Therefore, the gamma-exponential method can be seen as a refinement of the inverse square root method.

Monte Carlo experiments

In this section we discuss three Monte Carlo experiments, which we have implemented with MATLAB. Both employ the Monte Carlo simulation approach discussed in Appendix A. In these Monte Carlo experiments, 1000 samples were created and analyzed for each sample size. Each Monte Carlo experiment was conducted at least twice; i.e., at least two instances of each Monte Carlo experiment were conducted, with all of the corresponding results compiled and summarized. The results obtained across different instances of the same Monte Carlo experiment were virtually identical. This consistency in the results was primarily due to the large number of samples (i.e., 1000) created and analyzed for each sample size.

The first Monte Carlo experiment builds on the results from MBLC's study to develop its population model. As such, the true path coefficients are rather strong; which, as will be seen, lead to small minimum sample size estimates. To illustrate the influence of the path coefficients' magnitudes on minimum sample size estimation, the second Monte Carlo experiment uses the paths of the model in MBLC's study, but with all path coefficients reduced by .25. The third Monte Carlo experiment has path coefficients varying from .1 to .35, thus further illustrating the performance of the various minimum sample size estimation methods under more extreme conditions (e.g., a very small path coefficient) than the two previous experiments. In these Monte Carlo experiments the PLS-SEM analyses employ the PLS Mode A algorithm and the bootstrapping method for P value estimation; these are discussed in appendices C and D.

First Monte Carlo experiment

Figures 3 to 5 show graphs relating power to sample size for each of the paths of the model in MBLC's study. The sample size points shown (i.e., 15, 20 … 50) were chosen to allow us to estimate the minimum required sample size, as well as to illustrate how the power values vary based on sample size and path coefficient magnitude.

Figure 3: Paths CI→CL (left, β_p=.506) and CE→CL (right, β_p=.536)
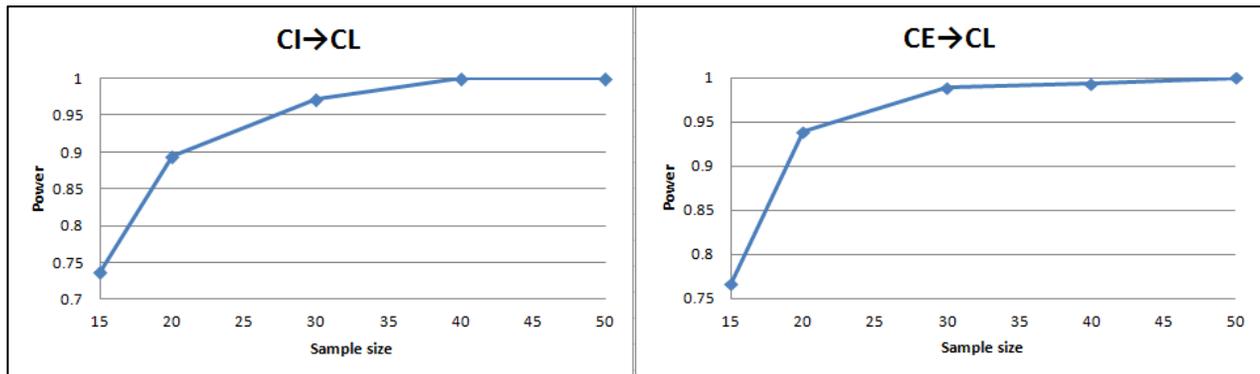
Figure 4: Paths CL→LO (left, β_p=.397) and CQ→LO (right, β_p=.525)
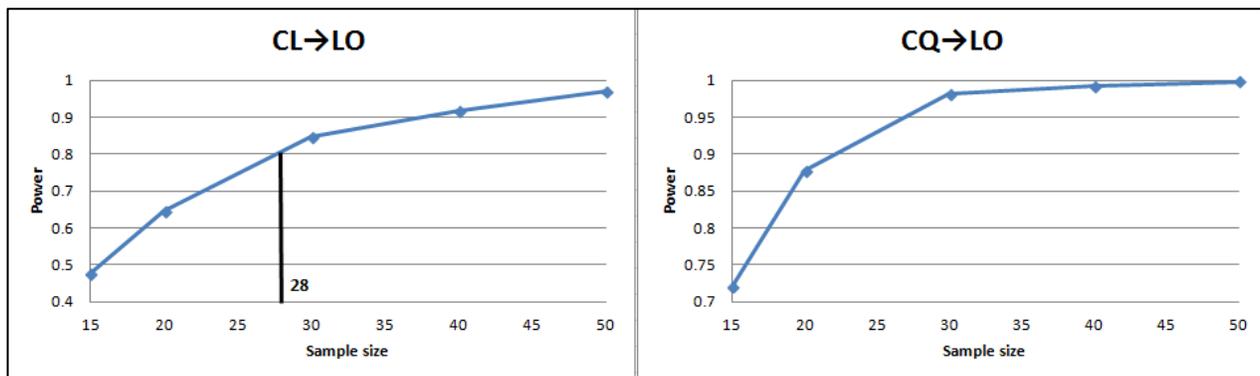
Figure 5: Paths CL→SO (left, β_p=.435) and CQ→SO (right, β_p=.498)
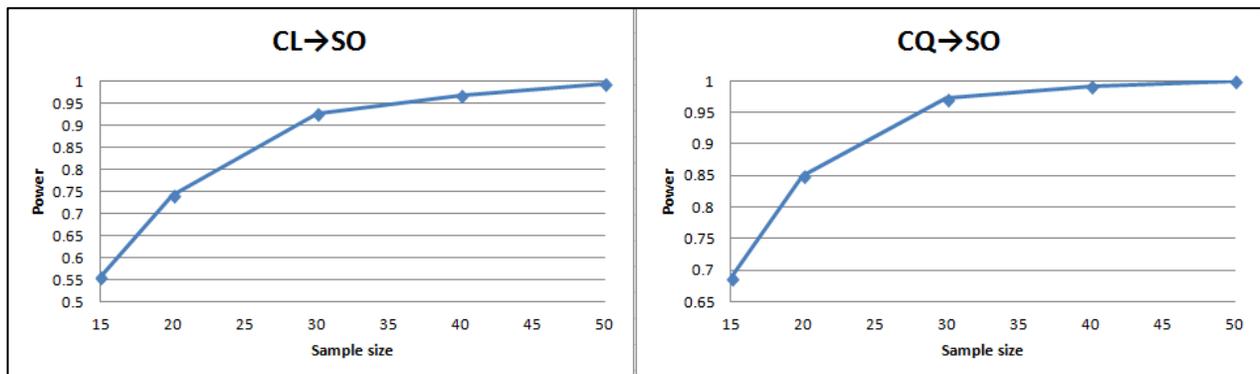
The population model had the same coefficients as those in MBLC's study's results. The power values shown are the percentages of path coefficients for which a significant effect was found, based on the significance level of .05 (i.e., P < .05) normally chosen in the field of IS. The P value calculation method employed was bootstrapping with 500 resamples.

As we can see, power values varied based on sample size and path coefficient magnitude. Power values increased as both sample sizes and path coefficient strengths increased. Therefore,

the minimum required sample size for the entire PLS-SEM analysis was the one at which the power of .8 was achieved for the path with the smallest magnitude, namely CL→LO (β_p=.397), indicated in the respective graph. Based on the graphs above, the minimum required sample size for the entire PLS-SEM analysis was estimated based on the Monte Carlo simulation to be 28.

Table 3 shows the estimates of the minimum required sample size based on the Monte Carlo simulation, the 10-times rule, as well as the R-squared, inverse square root and gamma-exponential methods. The Monte Carlo simulation estimate of 28 stands in for the true minimum required sample size. As we can see, the closest estimate to this true minimum required sample size is the gamma-exponential method estimate of 26.

Table 3: Performance of different estimation methods

Method	Minimum required sample size
Monte Carlo simulation	28
10-times rule	20
Minimum R-squared	33
Inverse square root	40
Gamma-exponential	26

Arguably all of the methods, except for the 10-times rule, led to minimum sample size estimates that would not lead to power levels drastically below the acceptable threshold. If used in an empirical study, the 10-times rule would lead to a sample size whose power would be approximately .65, which is well below the threshold of .8 for the .05 significance level. The gamma-exponential method would lead to a small underestimation: a sample size whose power would be just under .8 for the .05 significance level.

If used in an empirical study, the minimum R-squared and inverse square root methods would lead to relatively small and "harmless" overestimations of the true minimum required sample size needed. Arguably these overestimations would be harmless because they would lead to power values greater than the threshold of .8 for the .05 significance level – that is, more statistical power – without placing a significant demand on researchers for sample sizes much larger than necessary.

Second Monte Carlo experiment

Figures 6 to 8 show graphs relating power to sample size for each of the paths of the model in MBLC's study, but with all path coefficients reduced by .25. As with the first Monte Carlo experiment, the sample size points shown (i.e., 100, 200 … 400) were chosen to allow us to estimate the minimum required sample size, as well as to illustrate how the power values vary based on sample size and path coefficient magnitude.

Unlike the population model employed in the first Monte Carlo experiment, here the coefficients are much smaller than those in MBLC's study's results. Except for this key difference, the procedures to generate the numbers on the graphs are the same as those in the first Monte Carlo experiment. The power values shown are the percentages of path coefficients for which a significant effect was found, based on the significance level of .05. As before, P values were calculated through bootstrapping with 500 resamples.

Consistently with the first Monte Carlo experiment, power values varied based on sample size and path coefficient magnitude; increasing with both sample size and path coefficient magnitude. As expected, the minimum required sample size for the entire PLS-SEM analysis was the one at

which the power of .8 was achieved for the path with the smallest magnitude, namely CL→LO (β_p=.147). Based on the graphs above, the minimum required sample size for the entire PLS-SEM analysis was estimated to be 265 based on the Monte Carlo simulation.

Figure 6: Paths CI→CL (left, β_p=.256) and CE→CL (right, β_p=.286)
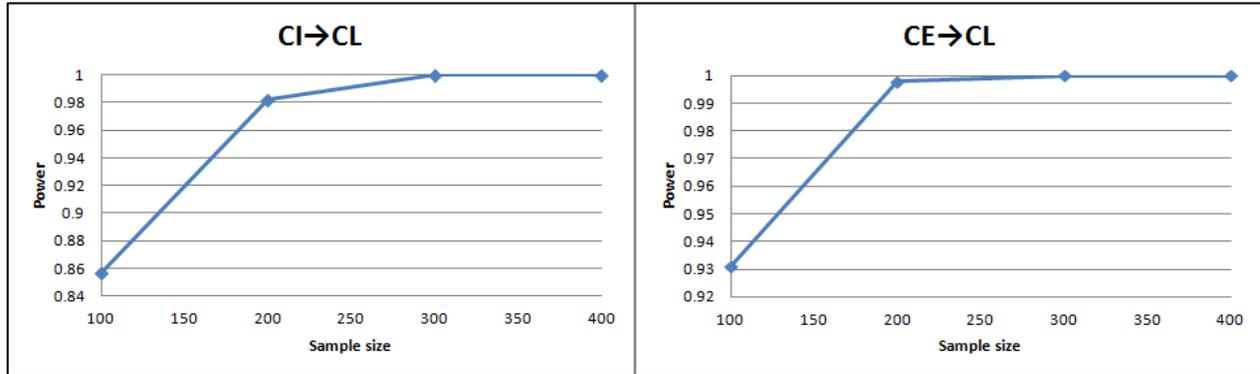
Figure 7: Paths CL→LO (left, β_p=.147) and CQ→LO (right, β_p=.275)
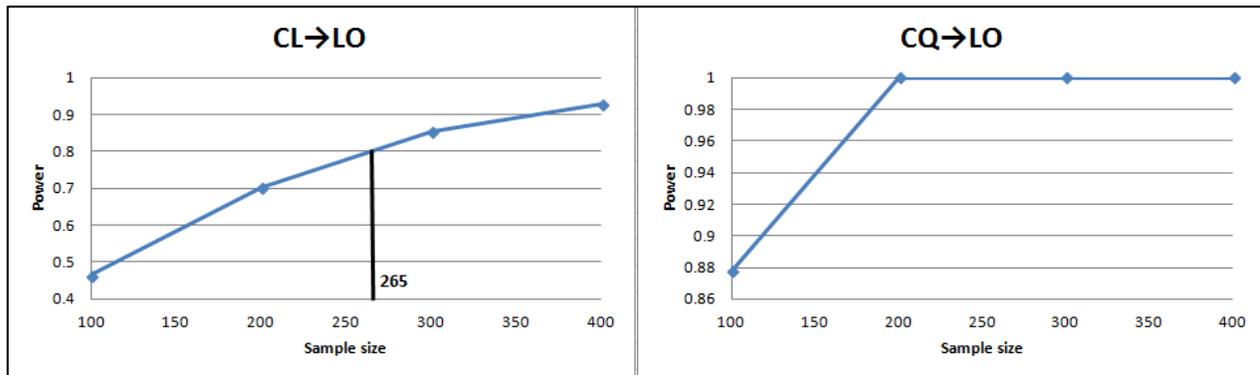
Figure 8: Paths CL→SO (left, β_p=.185) and CQ→SO (right, β_p=.248)

14
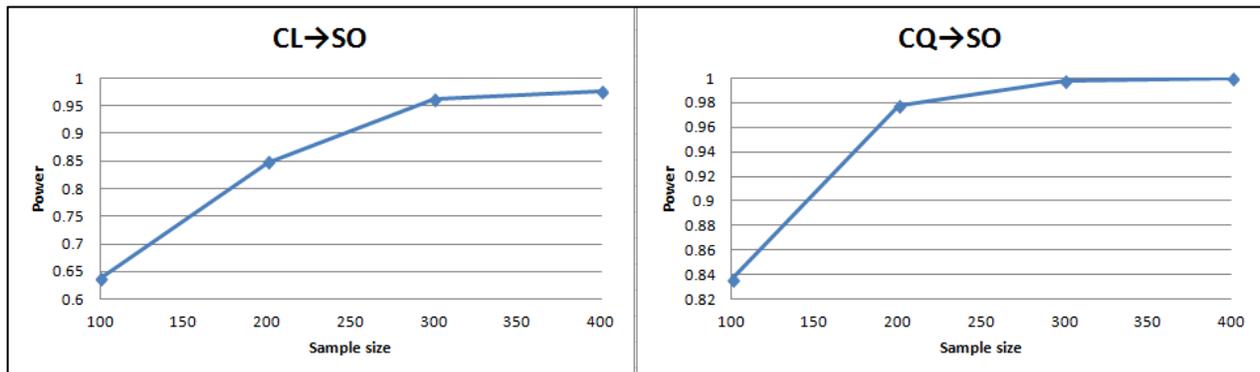
Table 4: Performance of different estimation methods

Method	Minimum required sample size
Monte Carlo simulation	265
10-times rule	20
Minimum R-squared	110
Inverse square root	287
Gamma-exponential	273

Table 4 shows the estimates of the minimum required sample size based on the Monte Carlo simulation, the 10-times rule, as well as the minimum R-squared, inverse square root and gamma-exponential methods. The minimum R-squared obtained in this second experiment was slightly lower than .1, which was used to produce the estimate employing the minimum R-squared method. As before, the Monte Carlo simulation estimate of 265 stands in for the true minimum required sample size. Again, as before, the closest estimate to this true minimum required sample size is the gamma-exponential method estimate of 273.

Here we can see that the 10-times rule and the R-squared method estimates were way off mark, significantly underestimating the true minimum required sample size. This was particularly true of the 10-times rule. If used in an empirical study, either of these methods would lead to sample sizes whose power would be well below the threshold of .8 for the .05 significance level.

When employed in an empirical study, the gamma-exponential and inverse square root methods would lead to relatively small and harmless overestimations of the true minimum required sample size needed. The overestimations would lead to power values greater than the threshold of .8 for the .05 significance level, arguably without placing a significant demand on researchers for sample sizes much larger than necessary. Even the estimate of 287, obtained via the inverse square root method, would require only 22 additional data points beyond the 265 necessary for the power threshold of .8 to be achieved.

Third Monte Carlo experiment

Figures 9 to 11 show graphs relating power to sample size for each of the paths of the model in MBLC's study, but with the path coefficients starting at the very small value of .1 and incrementally going up to .35. The sample size points shown (30, 50, 100 … 700) were chosen to allow us to estimate power values and the minimum required sample size under somewhat extreme conditions; starting with a very small path coefficient of .1 and with a very small sample size of 30. The sample size points were also chosen to illustrate how the power values incrementally grow based on path coefficient magnitude, as well as based on sample size.

We can see from the graphs, as expected from our mathematical reasoning presented earlier, that the path with the smallest magnitude CI→CL (β_p=.1) is the one that drove up the minimum required sample size. We also notice something interesting with this path, and also with the small path CE→CL (β_p=.15): a slightly anomalous behavior in the area involving the first two sample size points (i.e., 30 and 50). It seems that at these small sample sizes the power values for the two weakest paths are a bit higher than they should be. There reason for this is that the PLS Mode A algorithm we used in our PLS-SEM analyses tends to overestimate very weak paths at very small samples, by "capitalizing on error" (see, e.g., Goodhue et al., 2007).

Figure 9: Paths CI→CL (left, β_p=.100) and CE→CL (right, β_p=.150)
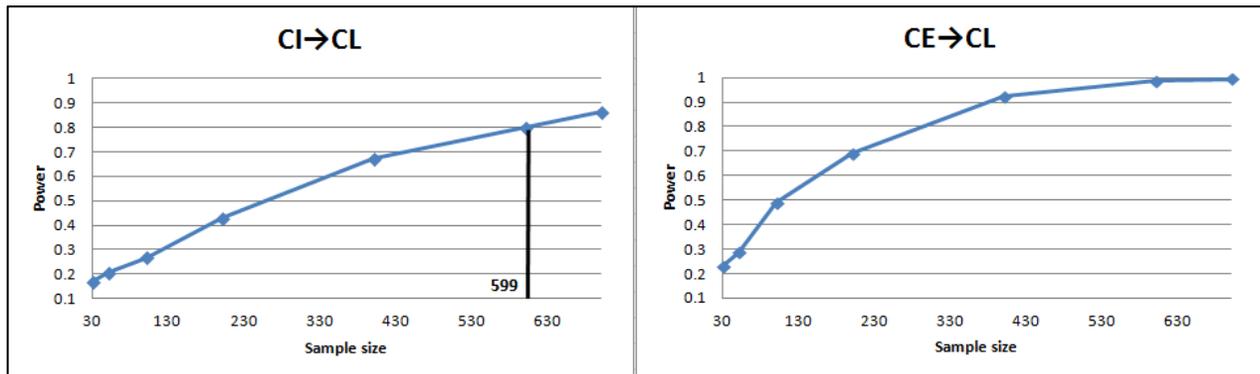
Figure 10: Paths CL→LO (left, β_p=.200) and CQ→LO (right, β_p=.250)
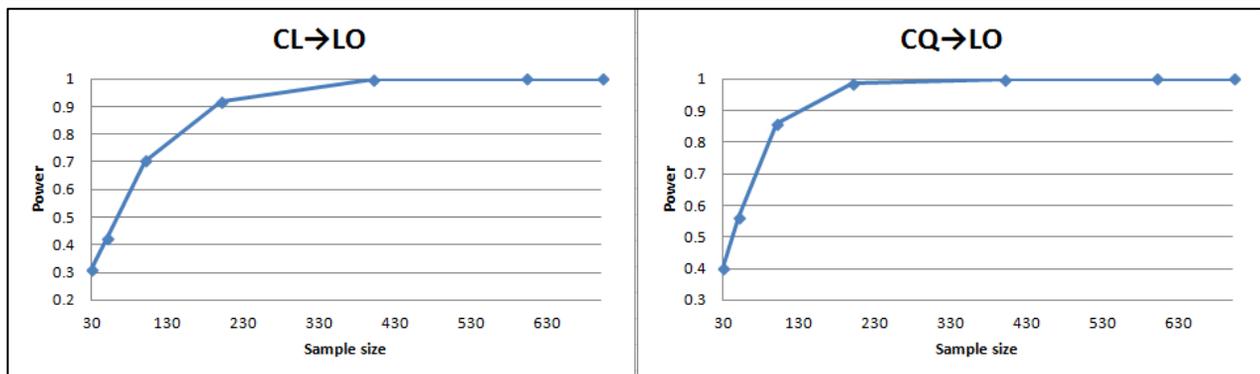
Figure 11: Paths CL→SO (left, β_p=.300) and CQ→SO (right, β_p=.350)
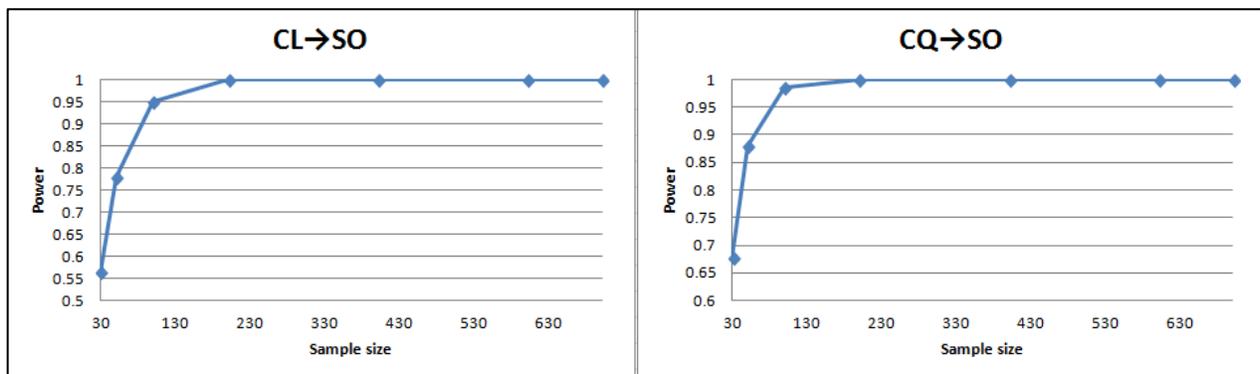
Capitalization on error is illustrated by the average path coefficient estimated in our Monte Carlo simulation for the path CI→CL (β_p=.1) at sample size 30, which was .108. This value is a bit higher than the true value of .1. This phenomenon may be at the very source of the mistaken belief that PLS-SEM has a remarkable ability to achieve acceptable power at very small sample sizes (Hair et al., 2011; 2014). Our results show that, even with capitalization on error, the power achieved for the small path CI→CL (β_p=.1) was too low at the small sample size of 30. It was

lower than .2, and thus well below the generally acceptable level of .8 for P < .05 (Cohen, 1988; 1992; Goodhue et al., 2012; Kock, 2016).

Table 5 shows the estimates of the minimum required sample size based on the various methods used in the two previous Monte Carlo experiments. The minimum R-squared obtained in this third experiment was much lower than .1, which was used to produce the estimate employing the minimum R-squared method. As with the two previous experiments, the Monte Carlo simulation estimate of 599 stands in for the true minimum required sample size. Again, as before, the closest estimate to this true minimum required sample size is the gamma-exponential method estimate of 605.

Table 5: Performance of different estimation methods

Method	Minimum required sample size
Monte Carlo simulation	599
10-times rule	20
Minimum R-squared	110
Inverse square root	619
Gamma-exponential	605

As with the second Monte Carlo experiment, we can see that the 10-times rule and the R-squared method estimates were again way off mark, significantly underestimating the true minimum required sample size. If used in an empirical study, either of these methods would lead to sample sizes whose power would be unacceptably low; well below the threshold of .8 for the .05 significance level. Even the sample size of 110 obtained through the minimum R-squared method would lead to a power level below .3; and this sample size is considerably greater than the sample size of 20 suggested by the 10-times rule. We can also see that the inverse square root and gamma-exponential methods yielded minimum sample size estimates that are fairly consistent with the one obtained via the Monte Carlo simulation method.

Additional Monte Carlo experiments

As part of our ongoing research on minimum required sample sizes in PLS-SEM, we have conducted a number of additional Monte Carlo experiments. These experiments included a variety of models, some simpler and others more complex than the model in MBLC's study. The results of these Monte Carlo experiments have been largely consistent with those of the three experiments presented above.

Several of the additional experiments that we have conducted included non-normal data. In some cases the level of non-normality was considerable; e.g., datasets created to have skewness and excess kurtosis values of 2.828 and 12 respectively. Consistently with past claims and related research (Chin, 1998; Hair et al., 2011; 2014; Kock, 2016), we found PLS-SEM to be fairly robust to deviations from normality, to the point that the Monte Carlo experiments yielded results that were virtually the same with both normal and non-normal data. Further analysis suggested that the underlying reasons for this are that least squares regression methods in general are quite robust to deviations from normality, and that so is bootstrapping. This is explained in more detail in Appendix E, with examples.

Discussion

In this section we discuss several issues in connection with minimum sample size estimation. We also provide several recommendations, primarily aimed at PLS-SEM users who are not methodological researchers. Among the issues addressed in this section are the method that arguably should be used for minimum sample size estimation, and minimum sample size estimation after and before data collection and analysis.

Which method should one use for minimum sample size estimation?

It is noteworthy that the gamma-exponential method, which is supposed to improve upon the precision of the inverse square root method, appears to: (a) slightly underestimate the minimum require sample size for small samples (e.g., $15 \geq N < 50$); and (b) slightly overestimate it for larger samples (e.g., $N \geq 100$). Does this mean that the gamma-exponential method yields incorrect estimates?

To answer this question we have also calculated the actual standard errors, as the actual standard deviations of the path coefficients, in the Monte Carlo experiments. These were calculated in addition to the standard error estimates yielded by bootstrapping and the gamma-exponential method corrections. As it turns out, the standard error estimates yielded by the gamma-exponential method corrections were closer to the actual values than those generated by bootstrapping.

For example, for the path CL→LO in the first Monte Carlo experiment (β_p=.397) at $N = 15$ the actual standard error was .181, the bootstrapping estimate was .225, and the gamma-exponential method estimate was .2. For the same path CL→LO in the second Monte Carlo experiment (β_p=.147) at $N = 100$ the actual standard error was .096, the bootstrapping estimate was .093, and the gamma-exponential method estimate was .096. These are consistent with the patterns of apparent imprecision that we observed, namely the slight underestimations and overestimations at different sample sizes.

In other words, the gamma-exponential method seems to yield the most precise estimates of standard errors, and thus minimum sample size estimates that are closest to the true values than any of the other methods discussed in this paper. This applies even to the Monte Carlo simulation method, because our implementation of this method relied on bootstrapping, which is the standard for PLS-SEM. This implementation decision was not made by mistake, as we attempted to mimic actual analyses conducted by empirical researchers employing PLS-SEM. It would be impossible to estimate actual standard errors in empirical PLS-SEM studies, since the true population values are not known.

Nevertheless, the gamma-exponential method is much more complex in its application (relying on a computer program) than the inverse square root method. The latter, in addition to being simpler (relying on a simple equation), is also fairly precise, leading to small overestimations that place light demands on researchers in terms of data points over the true minimum sample sizes required. Furthermore, the inverse square root method is "safe" in its slight imprecision, as it seems to always lead to small overestimations of the minimum sample sizes required.

Given the above, our recommendation for PLS-SEM users who are not methodological researchers is that they use the inverse square root method for minimum sample size estimation. They will be generating estimates that are both fairly precise and safe (slight overestimations),

with both normal and non-normal data. Their estimates will always be a little larger than the true minimum sample sizes required, but not by much.

Minimum sample size estimation after data collection and analysis

When minimum sample size estimation is conducted after data collection and analysis, its results can be used as a basis for additional data collection, as well as adjustments in the analysis and in the hypothesis testing assumptions. Minimum sample size estimation after data collection and analysis is known as retrospective estimation; as opposed to prospective estimation, conducted before data collection and analysis. Although there is debate on this topic, the latter (prospective) approach is generally recommended (Gerard et al., 1998; Nakagawa & Foster, 2004).

Additional data collection involves not only collecting additional data points, but also re-testing the model with the new dataset to ensure that the path coefficient with the minimum absolute magnitude has not decreased. Let us assume that a researcher collects 100 data points to test a PLS-SEM model, and finds that the path coefficient with the minimum absolute magnitude in the model is .237. Using the inverse square root method, the minimum required sample size is estimated to be 111. The researcher then proceeds to collect 11 additional data points and re-tests the model. If the path coefficient with the minimum absolute magnitude in the model is still .237 or higher, then the minimum sample size requirement is met.

Instead of collecting additional data points, the researcher may rely on adjustments in the analysis and in the hypothesis testing assumptions. Taking the example above as a basis, instead of collecting 11 additional data points the researcher may simplify the model somewhat by removing one or more competing links (i.e., links from multiple predictors to one criterion latent variable), particularly links competing with the link (or path) whose coefficient is .237. Clearly this should be informed by theory and past research; otherwise the empirical study becomes a data-fitting exercise.

Let us say that the researcher removed one link competing with the link whose path coefficient is .237. This removal would have a good chance of increasing the path coefficient with the minimum absolute magnitude, because each additional competing link tends to decrease the path coefficients for other competing links. Let us say that the path coefficient with the minimum absolute magnitude is .286 after the removal of one competing link. Using the inverse square root method, the minimum required sample size is estimated to be 76. This minimum required sample size is already met by the 100 data points originally collected. In this case, a simplification of the research model obviates the need for additional data collection.

An alternative to simplifying the model, which does not involve collecting more data either, is to regard a path coefficient that is too low in the context of a given sample size to be non-significant regardless of the corresponding P value. For example, let us assume that, with 100 data points, the two path coefficients with the smallest absolute magnitudes are .237 and .253, both found to be significant at P < .05 in an empirical study. Using the inverse square root method, the minimum required sample sizes associated with these two coefficients would respectively be 111 (as noted before) and 97. Here the researcher would regard the analysis to have failed to support the hypothesis associated with the .237 path, and succeeded in its support of the hypothesis associated with the .253 path.

Minimum sample size estimation before data collection and analysis

Minimum sample size estimation before data collection and analysis, or prospective estimation, is generally recommended over the retrospective approach of estimation after data collection and analysis (Gerard et al., 1998; Nakagawa & Foster, 2004). In prospective estimation, the researcher must decide at the outset the acceptable value of the path coefficient with the minimum absolute magnitude. This is likely to drive hypothesis testing beyond considerations regarding P values. In this context, an important question is: What is a reasonable acceptable value of the path coefficient with the minimum absolute magnitude in a model?

Based on Cohen's (1988; 1992) power assessment guidelines, a reasonable answer to this question would be a value that would satisfy $\beta^2/(1 - \beta^2) > .02$ in a very simple model with only one predictor and one criterion latent variable. In other words, the effect size measured via Cohen's f^2 coefficient in a model with only two variables X and Y, linked as $X{\rightarrow}Y$, would have to be greater than Cohen's (1988; 1992) minimum acceptable effect size of .02.

More complex models would tend to lead to lower effect sizes, because such models would likely include more competing links. Given this, we could set as our target an effect size that is twice Cohen's (1988; 1992) minimum acceptable, namely an effect size of .04. Our continuing research on this topic, including a variety of targeted Monte Carlo simulations, suggests that this rule of thumb covers the vast majority of models; including fairly complex models, as long as they are free of vertical and lateral collinearity (Kock & Lynn, 2012). The corresponding inequality for this proposed rule of thumb would be $\beta^2/(1 - \beta^2) > .04$, whose solution is $\beta \geq .197$.

Using the inverse square root method, the above would lead to a minimum required sample size of 160. Given this, another general rule of thumb could be proposed, this one as an answer to the following question: What is a reasonable value for minimum sample size, if we do not know in advance the value of the path coefficient with the minimum absolute magnitude? The answer would be 160, based on the inverse square root method. Based on the gamma-exponential method, the answer would be 146.

A different approach for prospective minimum sample size estimation is to set the acceptable value of the path coefficient with the minimum absolute magnitude based on past empirical research or the results of a pilot study. Either of these could suggest a large path coefficient of minimum absolute magnitude, which would lead to a relatively small sample size requirement. The danger here is in underestimating the minimum required sample size, which would call for conservative prospective estimations of the path coefficient of minimum absolute magnitude.

For example, if past empirical research or a pilot study suggests a path coefficient of minimum absolute magnitude of .35, the inverse square root method would yield a minimum required sample size of 51. Still, after having collected and analyzed 51 data points in an empirical study, a researcher would have to make sure that the path coefficient of minimum absolute magnitude was not lower than the expected .35. (A path coefficient of minimum absolute magnitude equal to or higher than .35 would have been acceptable.) If the path coefficient of minimum absolute magnitude turned out to be lower than the expected .35, the researcher would have to rely on approaches similar to those discussed earlier in connection with retrospective estimation (e.g., additional data collection).

Additional issues

The results of our third Monte Carlo experiment illustrated the phenomenon of capitalization on error, whereby a small path coefficient of .1 was overestimated by the PLS Mode A algorithm

we used in our PLS-SEM analyses. This clearly occurred for the sample size of 30, and seems to be a reflection of a general pattern that occurs with the PLS Mode A algorithm under certain conditions. Notable among those conditions are: (a) very small path coefficients (e.g., .1), and (b) very small sample sizes (e.g., 30).

Our results suggest that the methods we propose here for minimum sample size estimation are not affected by capitalization on error. The main reason for this is that our proposed methods tend to generate minimum sample size estimates for small path coefficients that are far above the sample sizes at which capitalization on error occurs. The fact that our methods focus on high power values (i.e., greater than .8) for minimum sample size estimation is a key element in avoiding bias due to capitalization on error. If we had tried to develop methods to estimate minimum sample sizes for low power values (e.g., .2), capitalization on error might become an issue.

Collecting and analyzing data at multiple levels of analysis can have an impact on minimum sample size estimation; e.g., collecting data at the individual and team levels. We see this in MBLC's study, where data from 17 project teams comprising 68 developers and 17 clients was collected, and where each group had 4 developers and 1 client. MBLC collected and analyzed data at the team level of analysis, which led to a rather small sample size of 17, and team aggregation of the individual data. If they had considered the individual team member to be the unit of analysis, the sample size would have been 85, but the analysis would have to become more complex in order to control for the effect of team membership on various hypothesized relationships. Another aspect that this type of analysis would arguably have to control for is whether the individual is a developer or a client. The topic of multi-level data collection and analysis is beyond the scope of the discussion presented here. Nevertheless, we provide in Appendix F a basic discussion of this topic in the context of MBLC's study.

Conclusion

IS researchers have been at the forefront of the development, software implementation, and use of PLS-SEM (Aguirre-Urreta & Marakas, 2013; Chin, 1998; Chin et al., 2003; Kock, 2010). One of the most fundamental issues in PLS-SEM is that of minimum sample size estimation, where the "10-times rule" method has been a favorite (Hair et al., 2011) due to its simplicity of application – it builds on the rule that the sample size should be greater than 10 times the maximum number of inner or outer model links pointing at any latent variable in the model.

In spite of the 10-times rule method's simplicity of application, it has been shown in the past to lead to inaccurate estimates (Goodhue et al., 2012). We proposed two related methods, based on mathematical equations, as alternatives for minimum sample size estimation in PLS-SEM: the inverse square root method, and the gamma-exponential method. Based on three Monte Carlo experiments, we demonstrated that both methods are fairly accurate. We also showed that the first method is particularly attractive in terms of its simplicity of application.

As demonstrated through our analyses and related discussion, the gamma-exponential method is much more complex in its application (relying on a computer program) than the inverse square root method. The latter, in addition to being simpler (relying on a simple equation), is also fairly precise, leading to small overestimations; and "safe" in its slight imprecision, apparently always leading to small overestimations of the minimum sample sizes required.

Consistently with these findings, it is our recommendation for PLS-SEM users who are not methodological researchers that they use the inverse square root method for minimum sample size estimation at the early stages of their research design. By doing so, those researchers will

generate estimates that are both fairly precise and safe, with both normal and non-normal data. Our analyses suggest that their estimates will always be somewhat larger than the true minimum sample sizes required, but not by much, placing light demands on data collection beyond what would actually be needed.

   The field of IS brings together researchers with a wide variety of skills and interests; counting among them methodological researchers, software developers, and expert users of methodological tools. The multidisciplinary nature of the field is perhaps one of its defining characteristics, and may be one of the reasons why IS has become a reference for other fields (Baskerville & Myers, 2002; Grover et al., 2006). Not surprisingly, IS researchers have provided the impetus for the widespread use of PLS-SEM (Chin, 1998; Chin et al., 2003; Kock, 2010), and IS researchers have also been at the forefront of questioning some of the claims in connection with PLS-SEM (Aguirre-Urreta & Marakas, 2013; Goodhue et al., 2012). It is our hope that this paper will contribute to this tradition of scholarly debate.

References

Aguirre-Urreta, M.I., & Marakas, G.M. (2013). Research note—Partial least squares and models with formatively specified endogenous constructs: A cautionary note. *Information Systems Research*, 25(4), 761-778.

Baskerville, R.L., & Myers, M.D. (2002). Information systems as a reference discipline. *MIS Quarterly*, 26(1), 1-14.

Bradley, R.V., Pratt, R.M.E., Byrd, T.A., Outlay, C.N., & Wynn, Jr., D.E. (2012). Enterprise architecture, IT effectiveness and the mediating role of IT alignment in US hospitals. *Information Systems Journal*, 22(2), 97-127.

Chin, W.W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly*, 22(1), vii-xvi.

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), 189-218.

Chin, W.W., Thatcher, J.B., & Wright, R.T. (2012). Assessing common method bias: Problems with the ULMC technique. *MIS Quarterly*, 36(3), 1003-1019.

Chintagunta, P. K. (2001). Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science*, 20(4), 442-456.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American*, 249(1), 116-130.

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297-316.

Efron, B., Rogosa, D., & Tibshirani, R. (2004). Resampling methods of estimation. In N.J. Smelser, & P.B. Baltes (Eds.). *International Encyclopedia of the Social & Behavioral Sciences (pp. 13216-13220)*. New York, NY: Elsevier.

Gerard, P.D., Smith, D.R., & Weerakkody, G. (1998). Limits of retrospective power analysis. *Journal of Wildlife Management*, 62(2), 801-807.

Goodhue, D., Lewis, W., & Thompson, R. (2007). Statistical power in analyzing interaction effects: Questioning the advantage of PLS with product indicators. *Information Systems Research*, 18(2), 211-227.

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

Grilli, L., & Rampichini, C. (2011). The role of sample cluster means in multilevel models. *Methodology*, 7(4), 121-133.

Grover, V., Gokhale, R., Lim, J., Coffey, J., & Ayyagari, R. (2006). A citation analysis of the evolution and state of information systems within a constellation of reference disciplines. *Journal of the Association for Information Systems*, 7(5), 270-325.

Gurland, J., & Tripathi R.C. (1971). A simple approximation for unbiased estimation of the standard deviation. *American Statistician*, 25(4), 30–32.

Hair, J.F., Hult, G.T.M., Ringle, C.M., & Sartedt, M. (2014). *A primer on partial least squares structural equation modeling*. Thousand Oaks, CA: Sage.

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.

Hardin, A.M., Fuller, M.A., & Valacich, J.S. (2006). Measuring group efficacy in virtual teams new questions in an old debate. *Small Group Research*, 37(1), 65-85.

Jak, S., Oort, F.J., & Dolan, C.V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling*, 20(2), 265-282.

Kipnis, C., & Varadhan, S.R.S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1), 1-19.

Klein, K.J., & Kozlowski, S.W. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, 3(3), 211-236.

Kline, R.B. (1998). *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. (2014a). *Stable P value calculation methods in PLS-SEM*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2014b). 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. (2015). One-tailed or two-tailed P values in PLS-SEM? *International Journal of e-Collaboration*, 11(2), 1-7.

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.

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg, Germany: Physica-Verlag.

Majchrzak, A., Beath, C.M., Lim, R.A., & Chin, W.W. (2005). Managing client dialogues during information systems design to facilitate client learning. *MIS Quarterly*, 29(4), 653-672.

Mandal, P., Mukhopadhyay, S., Bagchi, K., & Gunasekaran, A. (2012). The impact of organisational strategy, culture, people and technology management on organisational practice and performance: an empirical analysis. *International Journal of Information Systems and Change Management*, 6(2), 160-176.

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, 32(4), 355-373.

Miller, R.B., & Wichern, D.W. (1977). *Intermediate business statistics: Analysis of variance, regression and time series*. New York, NY: Holt, Rihehart and Winston.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599-620.

Nakagawa, S., & Foster, T.M. (2004). The case against retrospective statistical power analyses with an introduction to power analysis. *Acta Ethologica*, 7(2), 103-108.

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.

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

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209-233.

Rasoolimanesh, S. M., Dahalan, N., & Jaafar, M. (2016). Tourists' perceived value and satisfaction in a community-based homestay in the Lenggong Valley World Heritage Site. *Journal of Hospitality and Tourism Management*, 26(1), 72-81.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods*. New York, NY: Springer.

Schmiedel, T., vom Brocke, J., & Recker, J. (2014). Development and validation of an instrument to measure organizational cultures' support of business process management. *Information & Management*, 51(1), 43-56.

Schnabel, R.B., & Eskow, E. (1990). A new modified Cholesky factorization. *SIAM Journal on Scientific and Statistical Computing*, 11(6), 1136-1158.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science*, 44(4), 571-585.

Snijders, T., & Bosker, R. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. London, England: Sage.

van Mierlo, H., Vermunt, J. K., & Rutte, C. G. (2009). Composing group-level constructs from individual-level survey data. *Organizational Research Methods*, 12(2), 368-392.

Weakliem, D. L. (2016). *Hypothesis testing and model selection in the social sciences*. New York, NY: Guilford Publications.

Wold, H. (1980). Model construction and evaluation when theoretical knowledge is scarce. In J. Kmenta and J. B. Ramsey (Eds.), *Evaluation of econometric models* (pp. 47-74). Waltham, MA: Academic Press.

Appendix A: Monte Carlo simulations

In a Monte Carlo simulation the samples generated for each sample size point are based on a population model defined by the researcher, and build on common factor model assumptions, whose basic mathematical underpinnings are discussed in this appendix. For simplicity, and without any impact on the generality of the discussion presented here, we assume that variables are standardized – i.e., scaled to have a mean of zero and a standard deviation of 1.

Let ζ_i be the error variable that accounts for the variance in an endogenous latent variable F_i that is not explained by the predictor latent variables that point at F_i. Let F_j be one of the N_i predictor latent variables that point at an endogenous latent variable F_i. And let θ_{ij} be the standardized error variable that accounts for the variance in the indicator x_{ij} that is not explained by its latent variable F_i.

In a Monte Carlo simulation where multiple replications of a model are created (e.g., 1000 replications, or samples), error variables and exogenous variables can be created according to equations (A.1) to (A.3). In these equations $Rndn(N)$ is a function that returns a different normal random variable each time it is invoked, as a vector with N elements (where N is the sample size), and $Stdz(\cdot)$ is a function that returns a standardized variable.

$$\zeta_i \leftarrow Stdz\big(Rndn(N)\big). \tag{A.1}$$
$$F_j \leftarrow Stdz\big(Rndn(N)\big). \tag{A.2}$$
$$\theta_{ij} \leftarrow Stdz\big(Rndn(N)\big). \tag{A.3}$$

This assumes that simulated samples that follow normal distributions are desired. To obtain non-normal samples, transformations based on the normally-distributed variables can be employed. For example, equations (A.4) to (A.6) transform the normal variables into corresponding non-normal variables that follow a χ^2 distribution with 1 degree of freedom; a distribution with theoretical skewness and excess kurtosis values of 2.828 and 12 respectively.

$$\zeta_i \leftarrow Stdz\big({\zeta_i}^2\big). \tag{A.4}$$
$$F_j \leftarrow Stdz\big({F_j}^2\big). \tag{A.5}$$
$$\theta_{ij} \leftarrow Stdz\big({\theta_{ij}}^2\big). \tag{A.6}$$

After the error variables and exogenous latent variables are created, endogenous latent variables are produced based on the true population path coefficients defined beforehand by the researcher. This is indicated in (A.7), where R_{ij} are the correlations among the linked latent variables. Finally, indicators are created based on the true population loadings based on (A.8).

$$F_i = \sum_{j=1}^{N_i} \beta_{ij} F_j + \left(\sqrt{\left(1 - \sum_{j=1}^{N_i} \beta_{ij} R_{ij}\right)} \right) \zeta_i. \tag{A.7}$$

$$x_{ij} = \lambda_{ij} F_i + \left(\sqrt{1 - {\lambda_{ij}}^2} \right) \theta_{ij}, j = 1 \dots n_i. \tag{A.8}$$

Normally a set of samples (e.g., 1000 samples) is generated through the above steps for each sample size, with sample sizes varying incrementally. Generally speaking, a larger set of samples created in connection with each sample size (e.g., 1000 instead of 100), will lead to more precise and replicable measures of statistical power obtained via a Monte Carlo simulation for that particular sample size.

The above discussion refers to one of two main Monte Carlo simulation approaches, whereby both latent variables and indicators are generated. This approach has been commonly used in methodological PLS-SEM investigations published in IS and statistics outlets (Chin et al., 2012; Goodhue et al., 2012; Kock, 2016).

In the other main Monte Carlo simulation approach only indicators are generated (see, e.g., Mattson, 1997). This latter approach is based on the Cholesky factorization technique (Schnabel & Eskow, 1990), and has the disadvantage of not giving methodological researchers access to latent variable scores. Such scores may be needed in the estimation of certain model coefficients such as vertical and lateral collinearity variance inflation factors (Kock & Lynn, 2012), and are useful in the generation of non-normal data (Goodhue et al., 2012; Kock, 2016).

Appendix B: Excel spreadsheet with Visual Basic code

Figure B.1 shows the Excel spreadsheet that we have developed to implement the gamma-exponential method, by combining gamma and exponential smoothing function corrections applied to the standard error estimation employed in the inverse square root method. This Excel spreadsheet also shows the estimate obtained via the inverse square root method, for completeness.

Figure B.1: Excel spreadsheet

Estimating minimum sample size: Change the values in the yellow cells to obtain the values in the green cells					
T	1.645	(T ratio associated with P value threshold chosen)	P	0.050	(P value threshold for significance)
z	0.842	(z score associated with power level chosen)	W	0.800	(Power level chosen)
bmin	0.397	(Minimum absolute beta coefficient in model)			
N(EXPS+GAMM)	26	(Minimum required sample size: Exponential smoothing and gamma function methods)			
N(SQRT)	40	(Minimum required sample size: Square root method)			

As noted before, the T ratio for a chosen P value threshold is calculated as the z-score associated with 1 minus the P value threshold (e.g., $T_{.05} = z_{.95}$). In Excel a z-score associated with any value x is obtained using the function NORMINV(x,0,1) or the function NORMSINV(x).

Exhibit B.1 shows the Visual Basic code associated with the spreadsheet. We have developed this code so that, whenever any of the cells in yellow in the Excel spreadsheet changes, new values for the minimum required sample sizes are calculated. The code employs cell labels (e.g., T, z etc.) instead of standard cell references (e.g., B3, B4 etc.) in the "Range" object calls. The cell labels used are indicated on the spreadsheet next to their respective values.

Exhibit B.1: Visual Basic code

```
Private Sub Worksheet_Change(ByVal Target As Range)

    If Not Intersect(Target, Range("bmin, P, W")) Is Nothing Then
        Calc_N_EXPSnGAMM
    End If

End Sub

Sub Calc_N_EXPSnGAMM()

    Dim T As Double
    Dim z As Double
    Dim b As Double
    Dim cN As Long

    T = Range("T").Value
    z = Range("z").Value
    b = Range("bmin").Value
    cN = 1

    Do While Abs(b) * cN ^ 0.5 * Exp((Exp(1) * Abs(b)) / cN ^ 0.5) <= T + z
```

```
      cN = cN + 1

   Loop

   If cN <= 10 Then

      cN = 1

      Do While Abs(b) * cN ^ 0.5 * ((cN - 1) / 2) ^ 0.5 * Exp(WorksheetFunction.GammaLn(cN / 2)) /
Exp(WorksheetFunction.GammaLn((cN + 1) / 2)) <= T + z

         cN = cN + 1

      Loop

   End If


   Range("N_EXPS_GAMM").Value = cN

End Sub
```

The Visual Basic code is split into two main functions: "Worksheet_Change" and
"Calc_N_EXPSnGAMM". The code under "Calc_N_EXPSnGAMM" implements only the
gamma-exponential method. The inverse square root method is implemented without any Visual
Basic code, simply through the formula: ROUNDUP(((T+z)/ABS(bmin))^2,0).

Appendix C: PLS Mode A

Various PLS-SEM algorithms have been developed based on Wold's (1980) original design, of which an extensive discussion is provided by Lohmöller (1989). By far the most widely used of these algorithms is PLS Mode A. This is an iterative algorithm where indicator weight estimates $\hat{w}_{ij}$ are initially set to 1, and latent variable estimates $\hat{F}_i$ are initialized with a standardized vector of the summed indicators. Then the values of $\hat{F}_i$ are re-estimated as

$$\hat{F}_i := Stdz\left(\sum_{j=1}^{A_i} \hat{v}_{ij} \hat{F}_j\right). \tag{C.1}$$

The step implemented via (C.1) is known as the "inside approximation", where $Stdz(\cdot)$ is a function that returns a standardized column vector with N rows (where N is the sample size), and A_i is the number of latent variables $\hat{F}_j$ ($j = 1 \dots A_i$) that are "neighbors" of the latent variable $\hat{F}_i$. Latent variables are referred to as neighbors when they are linked to one another by arrows, either by pointing at or being pointed at by neighbor latent variables.

The weights $\hat{v}_{ij}$ are referred to as the "inner weights" (Lohmöller, 1989), and are estimated via three main schemes: centroid, factorial, and path weighting. In the centroid scheme the inner weights are set according to (C.2), as the signs (-1 or $+1$) of the estimated correlations among neighbor latent variables. In the factorial scheme the inner weights are set according to (C.3), as the correlations among neighbor latent variables. In the path weighting scheme the inner weights are set according to (C.4), as the path coefficients or correlations among neighbor latent variables, depending on whether the arrows go in or out respectively.

$$\hat{v}_{ij} := Sign\left(\Sigma_{\hat{F}_i\hat{F}_j}\right). \tag{C.2}$$

$$\hat{v}_{ij} := \Sigma_{\hat{F}_i\hat{F}_j}. \tag{C.3}$$

$$\begin{cases} \hat{v}_{ij} := \hat{\beta}_{ij}, & \text{if } \hat{F}_j \text{ points at } \hat{F}_i, \\ \hat{v}_{ij} := \Sigma_{\hat{F}_i\hat{F}_j}, & \text{if } \hat{F}_i \text{ points at } \hat{F}_j. \end{cases} \tag{C.4}$$

Next the PLS Mode A algorithm proceeds by estimating what are known as the "outer weights"; which is done by solving (C.5) for $\hat{w}_{ij}$, where: N_F is the total number of latent variables in the model, n_i is the number of indicators associated with latent variable F_i, and $\hat{\epsilon}_{ij}$ is the indicator error for the latent variable's *j*th indicator. This step yields estimates of the loadings of the indicators on their respective latent variables.

$$x_{ij} = \hat{F}_i\hat{w}_{ij} + \hat{\epsilon}_{ij}, i = 1 \dots N_F, j = 1 \dots n_i. \tag{C.5}$$

The PLS Mode A algorithm subsequently proceeds by estimating the latent variables as indicated in (C.6), through the step known as "outside approximation".

$$\hat{F}_i := Stdz\left(\sum_{j=1}^{n_i} \hat{w}_{ij} x_{ij}\right). \tag{C.6}$$

The foregoing steps are conducted iteratively until the outer weights $\widehat{w}_{ij}$ change by less than a small fraction. Then path coefficients are estimated by solving (C.7) for $\hat{\beta}_{ij}$. This essentially entails the solution of multiple ordinary least square regressions in "mini-models" with $\hat{F}_j$ predictor latent variables ($j = 1 \dots N_i$) pointing at each $\hat{F}_i$ criterion latent variable.

$$\hat{F}_i = \sum_{j=1}^{N_i} \hat{\beta}_{ij} \hat{F}_j + \hat{\zeta}_i. \tag{C.7}$$

Here $\hat{\beta}_{ij}$ is the estimated standardized partial regression (a.k.a. path) coefficient for the criterion-predictor relationship between F_i and $\hat{F}_j$, N_i is the number of predictors pointing at F_i in the model, and $\hat{\zeta}_i$ is the structural residual accounting for the variance in $\hat{F}_i$ that is not explained by the estimates of the latent variables that point at it in the model.

Appendix D: Bootstrapping

Bootstrapping (Diaconis & Efron, 1983) is the most widely used method for standard error estimation in PLS-SEM. Through the bootstrapping method, a standard error is generated for each path coefficient in a model, and then typically used to generate a T ratio, by dividing the path coefficient by the standard error. Next a P value is obtained for the T ratio using the incomplete beta function or a table relating P values to T ratios.

In PLS-SEM bootstrapping is typically applied to an empirical dataset, with the goal of creating multiple samples based on that dataset. Let $\mathcal{S}$ be a set of samples created based on an empirical dataset, where each sample in $\mathcal{S}$ is built by taking rows at random and with replacement (i.e., the same row can be repeated) from the original dataset.

Each sample in $\mathcal{S}$ has the same size (i.e., number of rows) as the original dataset. The set of samples denoted by $\mathcal{S}$ is also known in PLS-SEM as the resample set. In practical applications, the size of this resample set, or number of samples in $\mathcal{S}$, often varies from 100 to 500. Let N_S denote the size of this resample set, or the number of samples in $\mathcal{S}$.

The standard error estimate denoted as $\hat{S}_\beta$, obtained via bootstrapping for a given path coefficient β, is calculated according to (D.1), where: β_i is the path coefficient estimate for sample i, and $\bar{\beta}$ is the mean path coefficient across all samples. This is essentially the standard deviation of β.

$$\hat{S}_\beta = \sqrt{\frac{1}{N_S} \sum_{i=1}^{N_S} (\beta_i - \bar{\beta})^2}. \tag{D.1}$$

The bootstrapping approach to estimation of standard errors in PLS-SEM can be seen as a type of Monte Carlo simulation approach (Robert & Casella, 2013). Mimicking the sample creation process typically seen in Monte Carlo simulations, in bootstrapping many samples are created for subsequent analysis. But bootstrapping should not be confused with the Monte Carlo simulation method discussed earlier; in fact, bootstrapping is used as part of the simulations used in the Monte Carlo experiments.

The key difference between the two methods is that normally in Monte Carlo simulations the samples are created based on a true population model. In bootstrapping, on the other hand, an empirical dataset is used as a basis, from which multiple samples are created by taking rows at random and with replacement from the empirical dataset.

Appendix E: Non-normal data

Figure E.1 shows two histograms. The one on the left is for a standardized normally distributed variable, where skewness (indicated as "Skew." below the histogram) and excess kurtosis (indicated as "Kurt.") both approach zero. The histogram on the right shows a variable that follows a non-normal distribution, namely a χ^2 distribution with 1 degree of freedom. The variable on the right was created based on a transformation applied to the variable on the left, by squaring and standardizing it. This variable is in fact severely non-normal, with skewness of 2.80 and excess kurtosis of 10.98.

Figure E.1: Histograms of normal and non-normal data
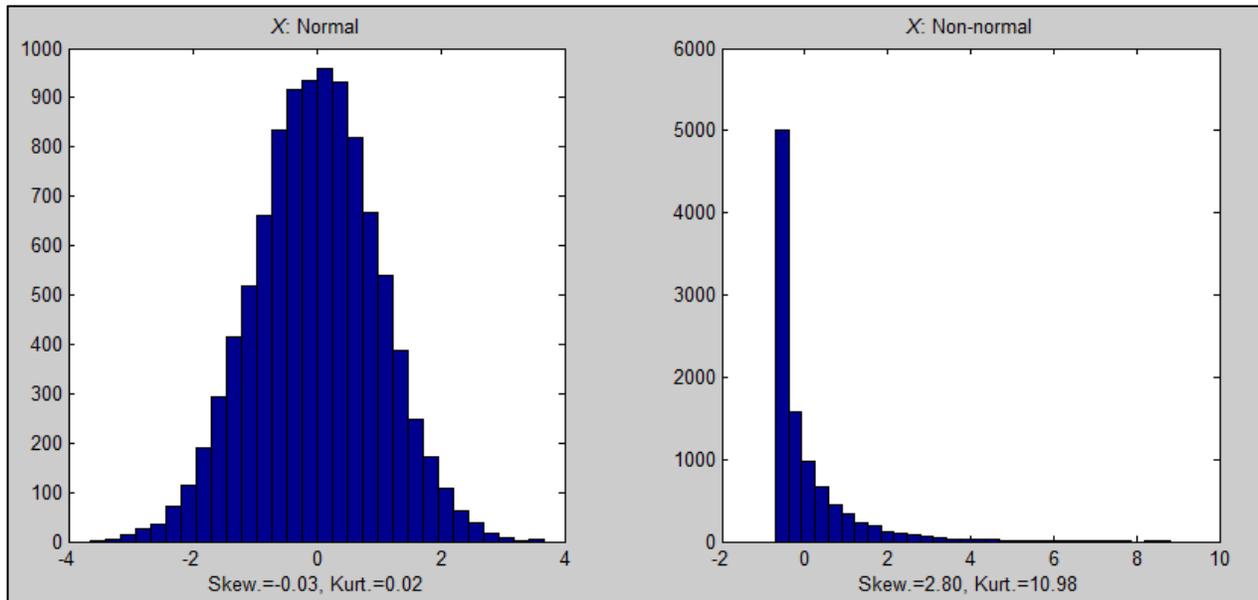
Figure E.2 shows data points and regression lines for three samples, where a predictor variable (X) is plotted on the horizontal axis and a criterion variable (Y) on the vertical axis. This figure is based on a simple model with only two variables, where $Y = \beta_{YX}X + \varepsilon$. This model's simplicity does not detract from its usefulness in our making our point regarding the robustness of least squares regression in the presence of non-normal data.

We also model the structural error (ε), to avoid non-normality propagation losses (Kock, 2016). In the figure we have the following: (left) both the predictor latent variable (X) and structural error (ε) are normal; (middle) the predictor (X) is non-normal but the structural error (ε) is normal; and (right) both the predictor (X) and structural error (ε) are non-normal. The data was created assuming the population model value of $\beta_{YX} = .3$. To minimize the bias due to sampling error, we used a sample size of 10000.

As can be inferred from the three graphs, the estimated least squares regression coefficient equals $.3$, which is the value of the corresponding population model coefficient, for each of the three cases. That is, even though the distribution of points shows a clear progression toward increasing non-normality as we move from left to right, we see no variation in the estimated least squares regression coefficients. In other words, we see no regression coefficient bias due to the different data distributions.

Figure E.2: Least squares regression coefficients with normal and non-normal data
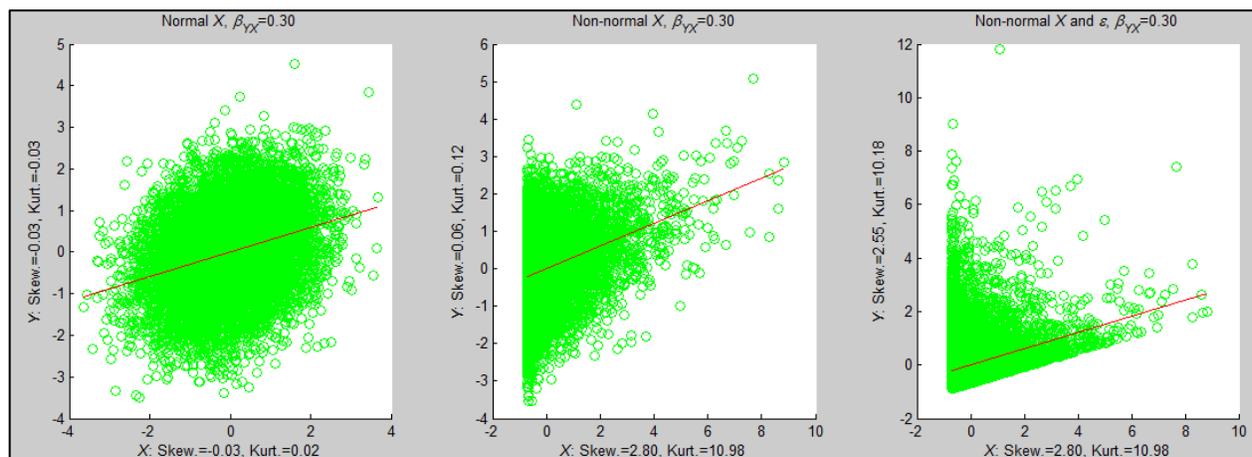
Notes: scales are standardized; left - predictor latent variable and error are normal; middle - predictor is non-normal but error is normal; right - predictor and error are non-normal.

Figure E.3 shows the distributions of least squares regression coefficients obtained via bootstrapping, in the form of histograms. The sample size here was set to 300, and the number of resamples to 500. The order of the graphs is the same as in the previous figure, with the same respective patterns of non-normality. Note that in all three histograms the values for skewness and excess kurtosis both approach zero, suggesting normality.

Figure E.3: Distributions of least squares regression coefficients obtained via bootstrapping
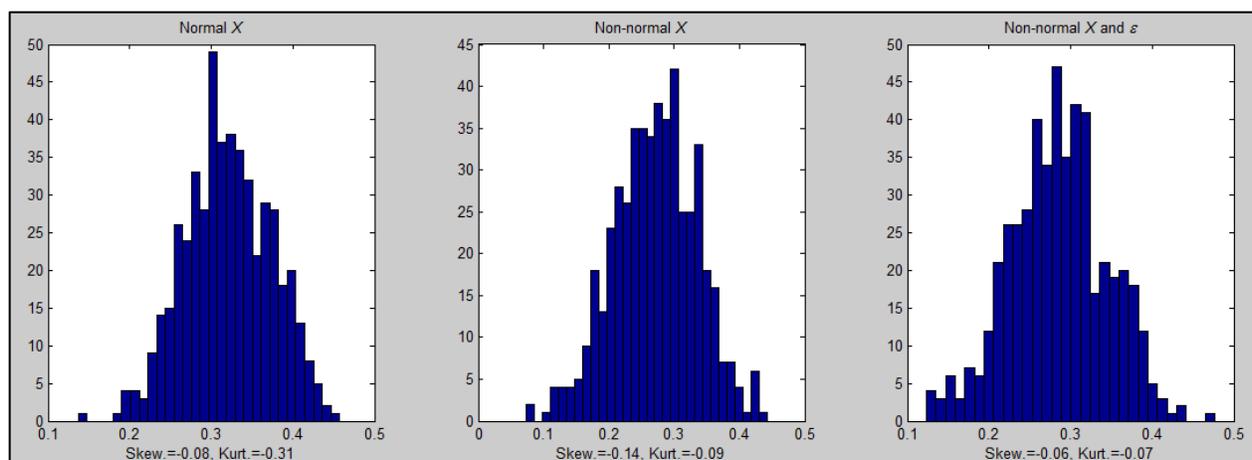
Notes: histograms show the distribution of least squares regression coefficients; the order of the graphs is the same as in the previous figure.

This figure demonstrates a remarkable property of bootstrapping, which is that it tends to yield normal distributions of estimates of the true least squares regression coefficient β_{YX} in variable associations of the type $Y = \beta_{YX}X + \varepsilon$ regardless of nature of the distributions of the variables Y, X and ε. Our continuing research on this topic, including a number of more complex simulations, leads us to conclude that this is generally true for more complex models as well, including

33

typical PLS-SEM models used in empirical IS studies. This property of bootstrapping together with the robustness of least squares regression coefficients in the presence of non-normality are, in our view, the main underlying reasons why PLS-SEM in general is quite robust to deviations from normality.

More research is needed to ascertain whether the above properties are retained in a recent development related to PLS-SEM: the consistent PLS technique (Dijkstra & Henseler, 2015). This new technique corrects certain parameters estimated via PLS Mode A employing the centroid scheme (Lohmöller, 1989); but does not generate latent variable scores or weights, which may be an obstacle to its widespread adoption without further developments aimed at consistently estimating those values. Those values are critical inputs for a number of tests now widely used in PLS-SEM, such as full collinearity and measurement invariance tests (Kock, 2014b; Kock & Lynn, 2012; Rasoolimanesh et al., 2016; Schmiedel et al., 2014).

Among the parameters corrected by consistent PLS are path coefficients, which tend to be underestimated by PLS Mode A and are used in our proposed minimum sample size estimation methods. However, it should be noted that in their practical example of the application of the consistent PLS technique to an empirical study in the field of IS, where they employed bootstrapping as we have done here, Dijkstra & Henseler (2015, p. 310) concluded that: "With regard to the two path coefficients, the differences between the estimates were rather small." This comment refers to path coefficient estimates generated via PLS Mode A and those corrected with the consistent PLS technique. The reason for the small difference is that psychometrically sound measures were used, leading to high reliabilities. Using psychometrically sound measures leading to high reliabilities is typically expected in well-executed empirical studies.

Appendix F: Multi-level data

Multi-level data is often found in the field of IS (see, e.g., Hardin et al., 2006) and it can have an impact on minimum sample size estimation. MBLC's study is an example of this. Data from 17 project teams comprising 68 developers and 17 clients was collected, where each group had 4 developers and 1 client. If we consider the team to be the unit of analysis, the sample size is 17. If we consider the individual team member to be the unit of analysis, the sample size is 85. Choosing the team as the unit of analysis requires some form of aggregation of team members' data at the team level (Klein & Kozlowski, 2000).

When multi-level data is available, a researcher may choose a level of analysis where there is no aggregation, such as the team member in MBLC's study, to obtain a larger sample. However, this leads to potential sources of bias, which must be addressed by the researcher. One alternative is to include additional variables that enable the researcher to control for the effects of group membership on each endogenous latent variable (Grilli & Rampichini, 2011). A frequently used option to implement this, which is not without problems (as discussed below), is to include variables storing the group means associated with existing variables (Grilli & Rampichini, 2011; Jak et al., 2013). This is illustrated in Figure F.1, for MBLC's study, where: TCI is a variable that stores the average cooperative interdependence (CI) in each team; TCE stores average collaborative elaboration (CE); TCL stores average client learning (CL); and TCQ stores average developers' communication quality (CQ).

Figure F.1: Including variables storing group means
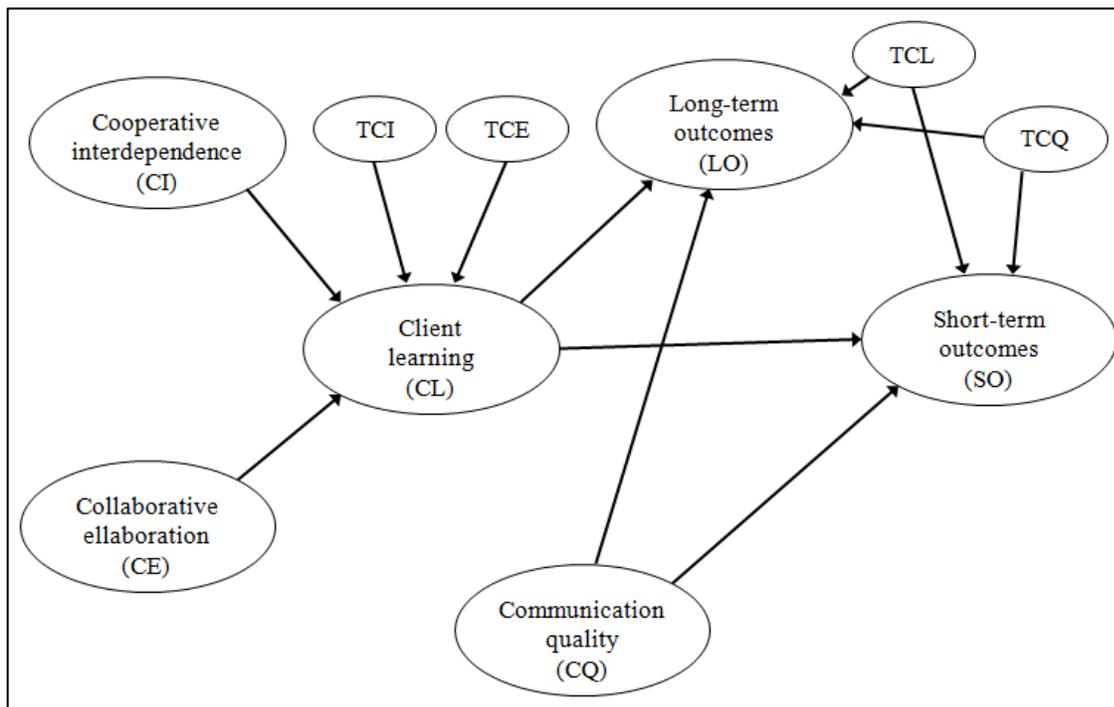
This approach is akin to that of including control variables into a model so that the results can be said to hold "regardless" of the effects of those variables; in this case, regardless of the influences arising from group membership. The inclusion of variables storing the group means

associated with existing variables would tend to decrease the path coefficients in the model, since they involve the insertion of competing links. An outcome of this would likely be a decrease in the path coefficient with the minimum absolute magnitude in the model. Nevertheless, such a decrease may not be enough to require a sample size greater than the one originally employed.

For example, let us consider a variation of MBLC's study where data at the team member level was used, yielding a sample size of 85. In this scenario, let us assume that no variables storing the group means were included, and that the path coefficient with the minimum absolute magnitude in the model was found to be .4. This is conservatively assumed to be only slightly higher than the coefficient of .397 for CL→LO obtained in the actual study; because there is more variation at the team member than at the team level of analysis, and thus less attenuation of path coefficients with respect to the true values (Kock, 2015; Lohmöller, 1989; Wold, 1980).

Let us now assume that the variables storing the group means – namely TCI, TCE, TCL and TCQ – are included in the model. The addition of TCL and TCQ alone is likely to bring down the path coefficient of minimum absolute magnitude in the model; let us assume it does, decreasing it from .4 to .3. This would be due to competing links being added to the model. Using the inverse square root method, the new path coefficient of minimum absolute magnitude in the model of .3 would lead to a minimum required sample size of 69. This requirement is met with a sample size of 85 at the team member level of analysis. That is, in this case it would have been advantageous, in terms of statistical power, to use a unit of analysis leading to a larger sample size; even though that decision would have made the model more complex due to the inclusion of additional variables storing the group means.

Being akin to controlling for the effect of demographic variables (e.g., controlling for the effect of gender), the approach of including variables storing group means is not without problems. A simple illustration where grouping is conducted demographically can help readers understand the approach more intuitively; and also help readers understand some of the approach's weaknesses. Let us assume that the 85 participants in MBLC's study were assigned to one of only two teams, one with only males and the other with females. Let us also assume that TCI, TCE, TCL and TCQ would each store two sets of different values for males and females. In this case, all of these four variables would end up being perfectly collinear with one another (i.e., absolute correlations of 1), and thus the model should only include one "group control" variable if collinearity were to be avoided. We could call this single group control variable GR (an acronym for gender); a variable with only two values (e.g., 1=female and 0=male). This new variable GR would replace the other four variables (i.e., TCI, TCE, TCL and TCQ) and point to all of the endogenous variables in the model: CL, LO and SO.

The illustration above highlights one of the problems with the approach of including variables storing group means: the fewer the number of groups available, the more likely it is that collinearity will occur. A high enough level of collinearity may significantly distort path coefficients, even if it is not perfect collinearity. For instance, with the original configuration in MBLC's study we might find, after a full collinearity test, that TCI and TCE are highly collinear (or redundant), and that so are TCL and TCQ; with full collinearity variance inflation factors above 5 for these variables (Kock & Lynn, 2012). To avert bias due to collinearity, at least one further step would have to be conducted. In this case, a researcher might use only one variable from each collinear pair; e.g., TCI in place of the pair TCI and TCE, and TCL in place of TCL and TCQ. Alternatively, the researcher might create two second-order variables (let us call them TCIE and TCLQ), and use only these second order latent variables in the model (see, e.g.,

Rasoolimanesh et al., 2016; Schmiedel et al., 2014). These second-order variables would respectively have as indicators: TCI and TCE, and TCL and TCQ.

Controlling for the effect of group membership by including variables storing group means may lead to another problem. Since it makes no assumptions about causal associations at the population level, it ignores the possibility of endogeneity (Chintagunta, 2001; Shaver, 1998) involving the variables storing group means. Let us consider the link TCI→CL for example. It allows us to estimate the path coefficient associated with the link CI→CL controlling for the effect of TCI, assuming that TCI and CI may be correlated. However, at the population level the correlation between TCI and CI may be due to the existence of a causal association instrumentally expressed by the link TCI→CI. If this is the case, TCI would also affect LO indirectly, through the network of links that connect these two variables. This would call for the inclusion of another direct link into the model: TCI→LO. Without this latter link, the path coefficient for the link CL→LO might be distorted by endogeneity; i.e., the correlation between TCI and the structural error term for the endogenous latent variable LO would not be accounted for.

Nevertheless, it is reasonable to assume that not including the direct link TCI→LO into the model would have a relatively minor biasing effect because the indirect effect of TCI on LO through the network of links that connect these two variables would likely be small. This is due to the fact that indirect effects in general tend to be small, because their magnitude is proportional to products of fractional coefficients, with each product significantly reducing the magnitude of the indirect effect.

The topic of multi-level data collection and analysis is complex and multi-faceted. The discussion presented here highlights key issues, and illustrates how multi-level data can influence minimum sample size estimation. Including variables storing group means is not without problems, as we have seen above. For broader discussions on the topic and alternative approaches, the reader is referred to Klein & Kozlowski (2000), Preacher et al. (2010), Raudenbush & Bryk (2002), Snijders & Bosker (1999), and van Mierlo et al. (2009).