One-tailed or two-tailed P values in PLS-SEM?

Ned Kock

Full reference:

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

Abstract

Should P values associated with path coefficients, as well as with other coefficients such as weights and loadings, be one-tailed or two-tailed? This question is answered in the context of structural equation modeling employing the partial least squares method (PLS-SEM), based on an illustrative model of the effect of e-collaboration technology use on job performance. A one-tailed test is recommended if the coefficient is assumed to have a sign (positive or negative), which should be reflected in the hypothesis that refers to the corresponding association. If no assumptions are made about coefficient sign, a two-tailed test is recommended. These recommendations apply to many other statistical methods that employ P values; including path analyses in general, with or without latent variables, plus univariate and multivariate regression analyses.

Keywords: E-collaboration; partial least squares; structural equation modeling; latent variable; indicator; one-tailed test; two-tailed test; Monte Carlo simulation.

Introduction

A common question often arises in the context of discussions about structural equation modeling (SEM) employing the partial least squares (PLS) method, referred to here as PLS-SEM (Kock, 2013b; 2014; Kock & Lynn, 2012), among researchers in the field of e-collaboration (Kock, 2005; Kock & Nosek, 2005) as well as many other fields. Should P values associated with path coefficients be one-tailed or two-tailed?

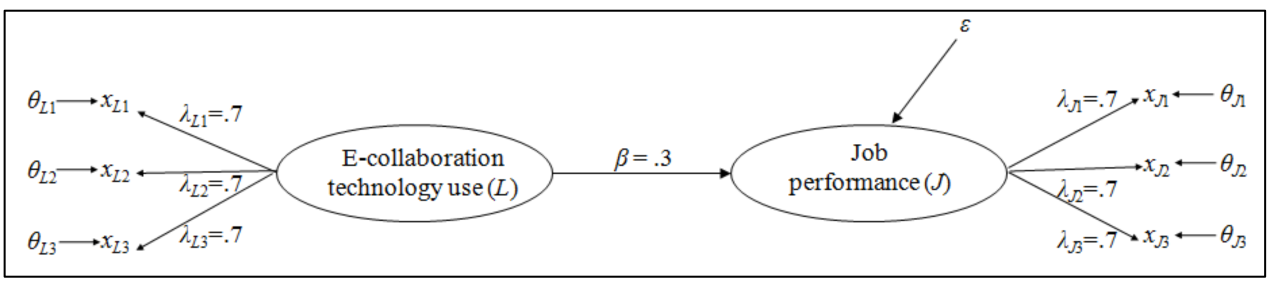
This is an important question because normally one-tailed tests yield lower P values than two-tailed tests. In fact, this is always the case when symmetrical distributions of path coefficients are assumed, such as Student's *t*-distributions. Therefore, the decision as to whether to use one-tailed or two-tailed tests can influence whether one or more hypotheses are accepted or rejected. This decision also influences the statistical power of a PLS-SEM analysis, with the power being higher with tests employing one-tailed P values.

We try to provide an answer to this question, which requires brief ancillary discussions of related topics – e.g., PLS-SEM's treatment of measurement error. While our discussion addresses path coefficients, it also applies to other coefficients such as weights and loadings. Even though the focus is on PLS-SEM, much of what is said here applies to many other statistical analysis techniques. Among these are path analyses in general, without or without latent variables, as well as univariate and multivariate regression analyses.

Illustrative model

The discussion presented in this study is based on the illustrative model shown in Figure 1. This model contains two latent variables, e-collaboration technology use (L) and job performance (J). Each latent variable is measured indirectly through three indicators.

Figure 1: Illustrative model



Let us assume that J, L, x_{Li} and x_{Ji} ($i = 1 \ldots 3$) are scaled to have a mean of zero and a standard deviation of one (i.e., these variables are standardized). Our illustrative model can then be described by equations (1), (2), and (3).

$$x_{Li} = \lambda_{Li} L + \theta_{Li},\ i = 1 \ldots 3. \quad (1)$$
$$x_{Ji} = \lambda_{Ji} J + \theta_{Ji},\ i = 1 \ldots 3. \quad (2)$$
$$J = \beta L + \varepsilon. \quad (3)$$

The path coefficient β and loadings λ_{Li} and λ_{Ji} ($i = 1 \ldots 3$) are assumed to describe the model at the population level, as true values. The population is made of teams of individuals who use an integrated e-collaboration technology including e-mail and voice conferencing to different degrees. That is, the unit of analysis is the team, not the individual.

The e-collaboration technology facilitates the work of the teams. Different values of job performance by the teams, where performance is evaluated by managers, are associated with different degrees of use of the e-collaboration technology.

PLS-SEM and measurement error

PLS-SEM algorithms estimate latent variable scores as exact linear combinations of their indicators (i.e., as "composites"). As such, they do not properly account for measurement error. This can be illustrated through (4) and (5); where latent variable scores are calculated properly accounting for, and not properly accounting for, the measurement error ϵ. Both equations denote the number of indicators as n.

$$F = \sum_{i=1}^{n} \omega_{Fi} x_{Fi} + \epsilon. \quad (4)$$

$$\hat{F} = \sum_{i=1}^{n} \widehat{\omega_{Fi}} x_{Fi}. \quad (5)$$

A full discussion of the effects of PLS-SEM not properly accounting for measurement error is outside the scope of this study. Nevertheless, one effect that will be noticed in the next section is that the path coefficient is attenuated, due to the correlation attenuation property (Nunnally & Bernstein, 1994) expressed in (6).

$$r(\widehat{F_i}, \widehat{F_j}) = r(F_i, F_j)\sqrt{\alpha_i \alpha_j}. \quad (6)$$

In this correlation attenuation equation, α_i and α_j denote the true reliabilities of the true latent variables F_i and F_j, which are estimated via PLS-SEM as $\widehat{F_i}$ and $\widehat{F_j}$. These true reliabilities can be estimated through the Cronbach's alpha coefficients for the latent variables.

Equation (7) expresses this general correlation attenuation equation in the more specific context of our illustrative model. In it, $\hat{J}$ and $\hat{L}$ are the PLS-SEM estimates of the true latent variables J and L.

$$r(\hat{J}, \hat{L}) = r(J, L)\sqrt{\alpha_i \alpha_j}. \quad (7)$$

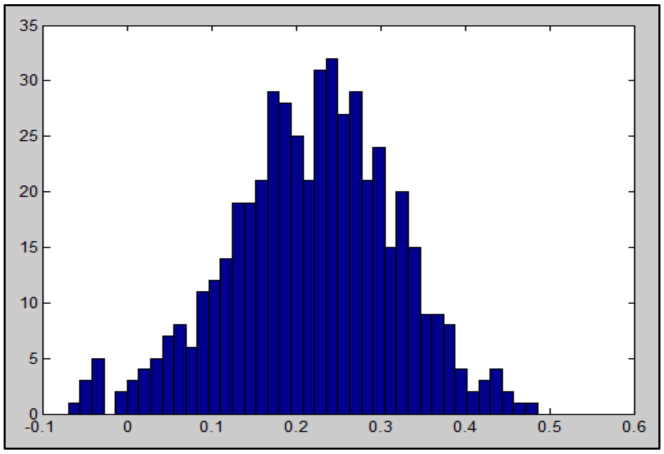
In our model the standardized path coefficient β is in fact equal to the true correlation $r(J, L)$, since the endogenous latent variable J has only one predictor (L). Even when this is not the case in more complex models, path coefficients tend to be attenuated in concert with their corresponding correlations.

Distribution of estimated coefficients across multiple samples

Figure 2 shows the distribution of values of the estimated path coefficient $\hat{\beta}$ across 500 samples of size 100. The samples were generated through a Monte Carlo simulation (Robert & Casella, 2005) based on the illustrative model. The data was created to follow normal distributions. Each sample was analyzed with the software WarpPLS, version 4.0 (Kock, 2013).

The analyses were conducted using the PLS Regression algorithm, which has been increasingly used in PLS-SEM (Guo et al., 2011; Kock, 2010).

Figure 2: Distribution of path coefficient estimates



Notes: N=100; PLS Regression algorithm used; values obtained through a Monte Carlo simulation with 500 samples (replications); shift to the left (from $.3$) in the distribution mean due to correlation attenuation.

As we can see, this distribution of values of the estimated path coefficient $\hat{\beta}$ across many samples does not appear to have a mean of $.3$, which is the true population mean. There appears to be a shift to the left. The reason for this is the correlation attenuation property discussed in the previous section; due to PLS-SEM algorithms in general, including PLS Regression, not properly accounting for measurement error.

The standard deviation of this distribution of values of the estimated path coefficient $\hat{\beta}$ across many samples is what is often referred to as the "standard error" associated with the estimate, denoted here as $\hat{\sigma}$. With the standard error $\hat{\sigma}$ and the mean estimated path coefficient $\bar{\beta}$ one can obtain the T ratio via (8).

$$T = \bar{\beta}/\hat{\sigma}. \tag{8}$$

The T ratio can then be used as a basis for the estimation of the one-tailed P value P_1 for $\hat{\beta}$ via integration through (9). In this equation $|T|$ is the absolute value of T and $F(t)$ is a function that refers to a Student's *t*-distribution.

$$P_1 = \int_{|T|}^{+\infty} F(t)dt. \tag{9}$$

Student's *t*-distributions are symmetrical about the mean. Therefore, the two-tailed P value P_2 for $\hat{\beta}$ can be obtained by multiplication of P_1 by 2, as indicated in (10).

$$P_2 = 2P_1. \tag{10}$$

Researchers employing PLS-SEM do not know the true population values of path coefficients and loadings prior to their analyses, and thus do not conduct Monte Carlo simulations to obtain estimates. They instead obtain estimates via resampling techniques, of which bootstrapping is the most widely used.

Resampling techniques can in fact be seen as part of a special class of Monte Carlo simulation techniques. They yield values for $\hat{\sigma}$ that approximate the true values, usually slightly underestimating them.

Also, in practice the value of P_1 is obtained via approaches other than integration, such as: specialized multivariate statistics and PLS-SEM software such as WarpPLS (which perform the integration themselves); general-purpose numeric calculation software such as R, MATLAB, and Excel; and published tables in statistics books and websites.

Using one-tailed and two-tailed P value estimations

Let us assume that we obtained the estimate $\hat{\beta} = .3$ for the path coefficient in our model. Do we use a one-tailed or two-tailed P value to estimate its significance? To answer this question we need to consider the hypothesis to which the estimate refers. The hypothesis is stated beforehand and incorporates the event whose complement's probability we are trying to ascertain via the test. Let us say that our hypothesis is as follows.

H_1: An increase in e-collaboration technology use (L) by a team is associated with an increase in job performance (J).

To test the significance of the estimate $\hat{\beta} = .3$ in the context of this hypothesis, via the calculation of a P value, is essentially to calculate the probability that the estimate $\hat{\beta} = .3$ is due to chance (the complement of what is stated in the hypothesis) given a set of pre-specified conditions. In this case, the set contains only one condition, which is that the path coefficient is positive, which is stated in the hypothesis.

If the effect is "real", and therefore not due to chance, the probability that it comes from a distribution that refers to no effect should be small. That is, this probability should be lower than a certain threshold, usually .05 (hence the oft-used $P < .05$ significance level). In PLS-SEM typically a distribution that refers to no effect is defined as a Student's *t*-distribution with a standard deviation that equals the standard error $\hat{\sigma}$ and that has a mean of zero.

The graph on the left in Figure 3 shows how this distribution would look like. The P value is calculated via integration. It is equal to the area indicated under the curve at the far right. Clearly the resulting probability refers to the one-tailed P value P_1 discussed in the previous section. This is the probability that the path coefficient estimate would be equal to or greater than $.3$.

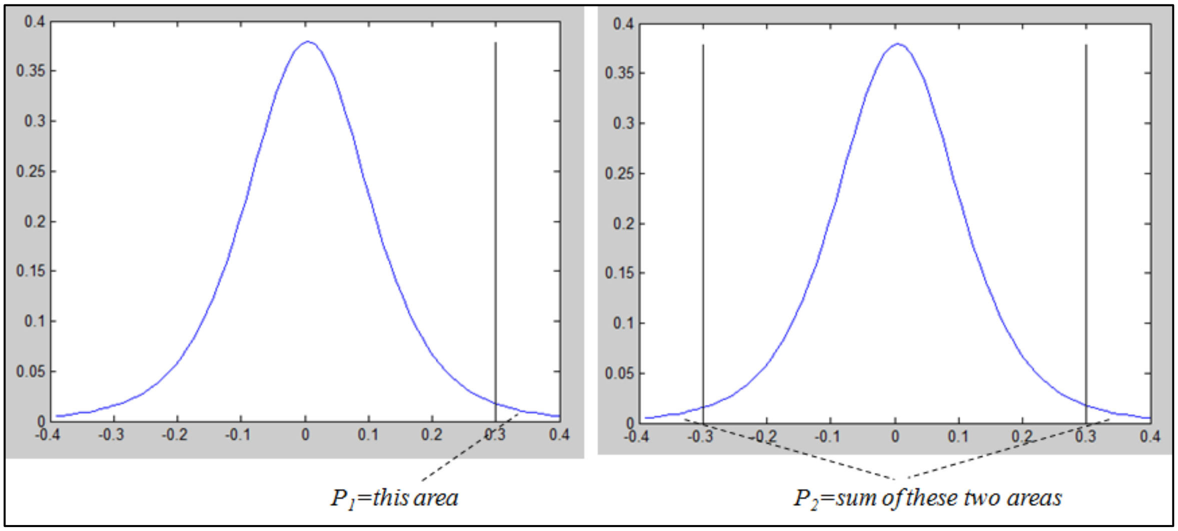
When would a two-tailed test be used? The answer, again, builds on the prior knowledge incorporated into the hypothesis being tested. Without the prior knowledge that the association between e-collaboration technology use (L) and job performance (J) is positive (i.e., that an increase in L leads to an increase in J), our hypothesis would likely be different. For example, it could be along the following lines:

H_2: There is an association between e-collaboration technology use (L) and job performance (J).

Here any value of $\hat{\beta}$ significantly lower or greater than zero would support the hypothesis. To test this hypothesis for a given estimate $\hat{\beta}$ obtained through a PLS-SEM analysis, we would again assume that a "no effect" path estimate would come from a Students' *t*-distribution with a mean of zero and with a standard deviation that equals the standard error $\hat{\sigma}$. But now we do not assume that the path coefficient is positive. Therefore we calculate the probability that we would obtain a path coefficient estimate that would be: equal to or greater than $\hat{\beta}$, or equal to or lower than $-\hat{\beta}$.

The graph on the right in Figure 3 shows how this probability would be calculated via integration as the sum of the areas indicated under the curve to the far left and far right. Clearly the resulting probability refers to the two-tailed P value P_2 discussed earlier.

Figure 3: One-tailed and two-tailed P value estimations



Notes: schematic representations; axes scales adjusted for illustration purposes.

Both graphs in Figure 3 are schematic representations, with the axes scales adjusted for illustration purposes. In the graphs of the actual distributions the areas used for P value estimation are often too small to be effectively used in visual illustrations of those areas under the probability distribution curves.

It is noteworthy that, in the discussion above, the hypothesized direction of causality of the effect ($L \rightarrow J$ or $J \rightarrow L$) is not as important in defining whether the test is one-tailed or two-tailed as the hypothesized sign of the effect. The hypothesized direction of causality could, under certain conditions, be important in defining the method of estimation of the path coefficient. This is particularly true if we assume that the relationship between the latent variables is nonlinear. In this case, the hypothesized direction of causality of the effect would become much more important.

Should the relationship be assumed to be nonlinear, thus leading to a nonlinear analysis (Kock, 2010; 2013), we would obtain different estimates for the nonlinear path estimate going in one direction $\hat{\beta}_{JL}$ and the other $\hat{\beta}_{LJ}$. This is an interesting property of nonlinear analyses that may have many useful applications. Among these applications is possibly that of causality assessment (Kock, 2013).

The meaning of the nonlinear path estimate would be different from that of the linear path estimate, since it would no longer refer to a fixed gradient, as a linear path estimate does. In the nonlinear case the gradients $\partial\hat{J}/\partial L$ and $\partial\hat{L}/\partial J$ would change for different values of the latent variables. This would have implications that arguably go beyond the scope of this study. Generally speaking, the sign of the nonlinear path estimate refers to the overall sign of the nonlinear relationship, or the sign of the "linear equivalent" of the nonlinear relationship.

Discussion and concluding remarks

The path attenuation phenomenon discussed earlier, stemming from PLS-SEM algorithms in general not properly accounting for measurement error, has an interesting influence on P value estimation using the approach discussed. It makes it more conservative. The reason is that the path coefficients estimated via PLS-SEM are closer to zero than the true path coefficients, which makes the area under the curve that refers to the P value normally greater than it would have been should an unbiased method be used. This leads to higher P values, other things being equal (e.g., the same resampling technique is used).

It may seem peculiar that prior knowledge incorporated into a hypothesis influences the test of the hypothesis. Nevertheless, this is consistent with the notion that, in frequentist inference, the conditional probability of any event is calculated based on a smaller set of possible events than the corresponding unconditional probability. This applies to events specified in hypotheses.

This leads to an interesting question. If our hypothesis incorporates the prior knowledge that $\beta > 0$ and our estimate turns out to violate this prior knowledge (e.g., $\hat{\beta} = -.3$), would a one-tailed test applied to $\hat{\beta} = -.3$ be acceptable? The answer is “no”, because if the prior knowledge incorporated into the hypothesis is not supported by the evidence (i.e., the negative path coefficient estimate), then the hypothesis is falsified outright. If a hypothesis incorporates the belief that $\beta > 0$ and we obtain an estimate $\hat{\beta} = -.3$ then the hypothesis is in fact falsified without the need for the calculation of a P value.

This highlights the fact that prior knowledge is important in the theorizing process that often precedes empirical research. Prior knowledge comes from thorough reviews of pertinent theories and past empirical research. The more prior knowledge is brought into empirical research, the more the research moves toward the confirmatory end of the exploratory-confirmatory spectrum. Generally speaking, bringing credible prior knowledge into empirical research is a “good thing”, and allows one to lower the threshold of evidence needed to ascertain the likelihood of an event that builds on that prior knowledge. Nevertheless, prior knowledge comes with an “inferential cost”, as discussed above.

Some researchers have suggested that P value estimation should be carried out directly from bootstrapping distributions. However, it should be clear that if we had used the distribution of path estimates obtained via bootstrapping in our tests instead of a Student’s *t*-distribution, a one-tailed estimation of P values would likely yield distorted results. This would have happened because bootstrapping distributions are usually asymmetrical, as our Monte Carlo-generated distribution was, with the degree of asymmetry varying depending on both data distributions and model characteristics.

We hope that the discussion presented here will help e-collaboration researchers who employ PLS-SEM, as well as researchers in other fields who use this multivariate analysis method, to decide whether to use one-tailed or two-tailed P values under different circumstances. Even though our discussion addresses primarily path coefficients, it also applies to other coefficients such as weights and loadings. While the discussion focuses on PLS-SEM, it applies to many other statistical analysis techniques. Among these are path analyses in general, with or without latent variables, as well as univariate and multivariate regression analyses.

Acknowledgments

The author is the developer of the software WarpPLS, which has over 7,000 users in more than 33 different countries at the time of this writing, and moderator of the PLS-SEM e-mail distribution list. He is grateful to those users, and to the members of the PLS-SEM e-mail distribution list, for questions, comments, and discussions on topics related to the use of WarpPLS.

References

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

Kock, N. (2005). What is e-collaboration. *International Journal of e-Collaboration*, 1(1), 1-7.

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. (2013). *WarpPLS 4.0 User Manual*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2013b). Using WarpPLS in e-collaboration studies: What if I have only one group and one condition? International Journal of e-Collaboration, 9(3), 1-12.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1-13.

Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.

Kock, N., & Nosek, J. (2005). Expanding the boundaries of e-collaboration. *IEEE Transactions on Professional Communication*, 48(1), 1-9.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Robert, C.P., & Casella, G. (2005). *Monte Carlo statistical methods*. New York, NY: Springer.