Why You Shouldn’t Use PLS:
Four Reasons to Be Uneasy About Using PLS In Analyzing Path Models

Dale L. Goodhue      Ron Thompson                 William Lewis
Terry College of Business   Schools of Business       Wake Forest University
University of Georgia          Wake Forest University
Dgoodhue@terry.gua.edu  thompsrl@wfu.edu
william.w.lewis@gmail.com

Abstract

When it was originally introduced, Partial Least Squares (PLS) was designed primarily for exploratory studies focusing on prediction (rather than hypothesis testing). Over time, however, PLS has become a very popular statistical analysis technique for testing hypothesized relationships (confirmatory studies) within MIS research. In this paper we note some challenges that have been raised relative to PLS, and then focus on four assertions concerning its use that we believe are problematic. We show that these frequently stated assertions of PLS strength actually suggest important weaknesses. In particular, we show evidence that PLS seems to capitalize on chance correlations among indicators at sample sizes typically used in MIS research. Our belief is that many PLS users are unaware of the trade-offs involved in PLS use, and we recommend a much reduced (or possibly non-existent) role for PLS in confirmatory work.

1. Introduction

PLS is now the preferred statistical technique for use in analyzing path models in the top MIS research journals [17]. However, there is an increasing body of research suggesting that PLS does not have all of the positive characteristics that are often ascribed to it. (We note that a distinction needs to be drawn between PLS path modeling, as implemented in software packages such as PLS-Graph and SmartPLS, and PLS regression, as implemented in packages such as WarpPLS. Our focus in this paper is on PLS path modeling, which has been by far the more popular approach in MIS research to date).

Much of the research challenging claims about PLS involves Monte Carlo simulations that test PLS’s performance in comparison to other techniques such as regression or covariance-based structural equation modeling (CB-SEM). In this paper we first provide a brief review of some of these empirical challenges to PLS, and then focus on four (more general) assertions that have been made about PLS and its strengths. Our goal is to examine those assertions more closely to understand exactly what is being claimed, and what evidence there is to support those claims. Three of these assertions are quite commonly stated, the fourth (number 3 below) somewhat less frequently. The assertions are the following:

(1) PLS presumes “no distributional form at all” [6, p. 442].
(2) PLS is especially appropriate for exploration and prediction versus confirmation and hypothesis testing, “where the phenomena in question are …. relatively new or changing, and where the theoretical model or measures are not well formed” [1, p. 333].
(3) Alternative PLS models may be compared using a pseudo F test. Mathieson et al. present a pseudo F statistic which allows a test of the statistical significance of the improvement in fit from the addition or deletion of a single path to a given PLS model. [12, p. 104].
(4) PLS estimates “are asymptotically correct in the joint sense of consistency (large number of cases) and consistency at large conditions (large number of indicators for each latent variable” [10, p. 266]).

We will show that all four of these supposedly positive assertions about PLS are actually indicators of potential weaknesses for its use in MIS research. In the next section we provide a brief review of some of the empirical challenges to PLS’s statistical efficacy, and then explore these four assertions in more detail.
2. Empirical Challenges to PLS

There have been a number of challenges to PLS’s sometimes touted advantages.

Advantage at Small Sample Size. Goodhue, Lewis and Thompson [8] have presented evidence challenging the assertion that PLS has advantages over CB-SEM when sample size is small. Instead, their Monte Carlo simulations suggest that regression, PLS and CB-SEM all lose significant amounts of statistical power with sample sizes smaller than 90 or so, because of increasingly large standard deviations. Further, all three techniques lose statistical power at about the same rate.

Advantage with non-normal data. Gefen, Rigdon and Straub [7] have suggested that this is no longer a reason to choose PLS over CB-SEM, since new CB-SEM estimation approaches have been developed that are robust in the face of non-normal data. In addition, even using the default maximum likelihood estimation approach in CB-SEM, Goodhue et al. [8] have shown that all three techniques (PLS, CB-SEM and regression) are remarkably robust to moderate departures from normality, and equally so. All also suffer a significant loss of power with extreme departures from normality, again about equally.

Advantage with formative constructs. Another reason that is often cited for using PLS is that it is easier to model formative measures with PLS than with CB-SEM approaches [2]. Some researchers have questioned the use of formative constructs with any statistical technique [4], while others have noted ways to address limitations in the use of CB-SEM with formative measures [7]. One study comparing the two approaches for modeling formative measures using Monte Carlo simulation produced somewhat mixed results [16], but the authors concluded that CB-SEM should probably be used instead of PLS unless the premises of CB-SEM (sample size and data distribution) were significantly violated.

Useful model fit statistics. PLS path modeling does not provide a single, overall test of the fit between the model and the data. While numerous heuristic criteria have been offered, none have been shown to consistently distinguish between correct and incorrect models. In a comprehensive Monte Carlo simulation test of different model fit criteria, Evermann and Tate [5, p. 16] found that “The criteria display a bewildering range of behavior, depending strongly, and in a complex way, on the model, type of misspecification, sample size, number of indicators, loadings, and strength of structural coefficients.” The authors concluded that none of the heuristic criteria commonly used for assessing model fit with PLS work effectively.

Advantages from minimizing the effect of measurement error. Some early descriptions of PLS have implied that the weighting of items in PLS minimizes the effect of measurement error [2]. Rönkkö and Ylitalo [18] used Monte Carlo simulation to test the effect of measurement error, and found that while the indicator weightings in PLS maximized the explained variance in the data, they also resulted in biased estimates and did not control for measurement error. They concluded that PLS was potentially a worse alternative than traditional path analysis [18, pg. 2].

We will now turn to the four non-empirical assertions mentioned in the introduction of this paper. In the process, we will introduce some new empirical results to help us explore one of the four assertions.

3. PLS has no distributional form at all

In contrast to regression or CB-SEM, PLS requires no distributional assumptions. While this sounds good on the surface, let’s look at what the assumptions for regression and CB-SEM actually are, and contrast that with the situation for PLS.

Regression Assumptions. For ordinary least squares (OLS) regression, the assumptions are the following:

i. The relationship between the Xs and Y is linear.
ii. The Xs are nonstochastic, and no exact linear relationship exists between any two Xs.
iii. The values of the error term:
   a. Have an expected value of zero and constant variance for all observations
   b. Are uncorrelated across observations
   c. Are normally distributed

When data in a regression meets these assumptions, then the parameter estimates can be shown mathematically to be optimal. More specifically, they are the best linear unbiased estimates possible (or BLUE, in the language of statisticians). Of all possible linear estimates, these have the smallest variance and the least bias. Of course we know that more often than not, actual data does not strictly meet these expectations. But many researchers have demonstrated that regression is remarkably robust to moderate deviations from these
characteristics. So for those rare datasets where all assumptions are met, regression estimates are optimal. Where the assumptions are nearly met, regression estimates are nearly optimal.

However, regression assumption ii is often not fully understood. “Nonstochastic” actually means that the values of the Xs are known with certainty (i.e., that the Xs are measured without error). So regression is mathematically optimal only when constructs are measured without error, which is not common in MIS research. When constructs are measured with error, Nunnally and Bernstein [14, pp. 241, 257] show that estimations of correlations between constructs (and thus of regression path estimates) are attenuated according to the following equation:

\[
\text{Apparent correlation}_{XY} = \frac{\text{Actual correlation}_{XY} \cdot \text{Sqrt}(\alpha_X \cdot \alpha_Y)}
\]

Here \(\alpha\) is Cronbach’s alpha, the reliability (of X or Y).

Translated into the situation for a regression with one X construct, and assuming that the reliability of X and of Y are the same, we see that the bias in a regression estimates is:

\[
\text{Reg. Bias} = \frac{\beta_{XY}}{\beta_{XY}} = \text{Sqrt}(\alpha_X \cdot \alpha_Y)
\]

(when \(\alpha_X = \alpha_Y\), Regression Bias = \(\alpha\))

So we have to be careful about saying that regression estimates are optimal when our constructs are measured with error. In fact they are biased by a very specific amount, based on the reliabilities of the construct measures.

**CB-SEM Assumptions.** For CB-SEM with maximum likelihood, the assumption is that all variables in the population are distributed according to a multivariate normal distribution. When this assumption is met, CB-SEM can be shown mathematically to produce estimates that are optimal. That is, they strictly minimize the difference between the correlation matrix implied by the solution and the sample correlation matrix. In other words, the parameter estimates are as close as can be achieved given the stated model and data.

Once again we know that most data does not conform to this assumption exactly, but it has been shown that CB-SEM is reasonably robust to moderate violations of this assumption. For those rare datasets where the assumptions are met, CB-SEM estimates are optimal. Where the assumptions are nearly met, CB-SEM estimates are nearly optimal.

In contrast to regression, this optimality condition includes the case where multiple indicators of a construct are all measured with error, as is typically the case in much MIS research. This is a potential advantage of CB-SEM over regression (and PLS).

**PLS Assumptions.** PLS has no presumption of any distribution assumptions. We could interpret this statement in one of two different ways. One of these is that PLS estimates could be shown mathematically to be optimal regardless of the distribution of the sample. The second way to interpret the statement is that PLS estimates are not expected to be optimal, and cannot be shown to be optimal, under any set of assumptions. While MIS researchers may think the first is true, it is the second that is asserted by the developers of PLS. Wold [21, p. 28] says “Sacrificing optimality, PLS rests content with consistency (Section 6.3), albeit in the qualified sense of consistency at large (Section 6.4)”. He goes on to say “the gains of sacrificing optimality are considerable: (i) explicit estimation of the case values of the LV’s; (ii) instant estimation; (iii) in developing the model by a dialog with the computer the investigator can at each step, rapidly and at low cost thanks to (i) and (ii), perform a variety of pilot explorations……Lost in the passage to multi-LV soft models is the overall optimization property.”

MIS researchers may think the gains mentioned by Wold are important or not, but there is no assertion of any characteristic of overall optimization here, regardless of what assumptions are met. To say that PLS does not have any distribution assumptions is in effect admitting that PLS estimates have not been mathematically shown to be optimal regardless of what assumptions are made.

**4. The pseudo F test can be used to compare competing PLS models.**

Although not mentioned as frequently as the other assertions dealt with here, the pseudo F test with PLS is used from time to time in our top journals [e.g. 17, 19, 22]. It is sometimes attributed to Cohen [3], but the earliest we found it presented in the context of PLS was in a paper by Mathieson, Peacock and Chin [12]. They start with an equation from Cohen [3] for effect size:

\[
f^2 = \frac{R_{\text{full}}^2 - R_{\text{excluded}}^2}{1 - R_{\text{full}}^2}
\]

They then suggest multiplying \(f^2\) by \((n-k-1)\) to give a “pseudo” \(F\) statistic with 1 and \(n-k\) degrees of freedom. This \(F\) statistic can be used to test whether
two nested models are different in a statistically significant way.

When regression is being used, this is indeed the F statistic explained by Cohen [3, pp. 408-411], though the particular version stated by Mathieson et al. presumes that only one predictor is being excluded, and there is no constant. However in regression, when only one predictor is being excluded, the above F test is mathematically equivalent to the t test on that predictor’s path, when that one predictor is included in the full regression [13, p. 281]. This is why, at least in regression, this F test is usually employed to test the impact of a collection of predictors rather than a single one. In regression if the F test were used with a single predictor, the statistical significance results would always be the same as for the t statistic. However, in PLS that is not true in general – the t test and the F test will give similar but not identical results. This is because in the first stage of the PLS algorithm when a given path between constructs is excluded, the various weights and loadings for the indicators will all shift somewhat, as will the resultant construct scores, making the comparison between the two models not quite equivalent.

We note that in his discussion of the F test, Cohen never uses the word pseudo, and never refers to PLS. He is discussing a true F statistic used with a true regression. The American Heritage dictionary [20] defines pseudo as “false or counterfeit; fake”. The true F statistic (in this context) is based on the values of the regression $R^2$, a well defined quantity with a precise mathematical meaning and precise mathematical properties. Given the way $R^2$ is determined in PLS, it does not have the same mathematical meaning, and thus using PLS’s $R^2$ in an F test comparison of two models is not strictly correct (in the sense of Cohen’s F test). Mathieson et al. are essentially acknowledging that this is the case by using the term “a pseudo F test.”

### 5. Exploration and Prediction versus Confirmation and Hypothesis Testing

Gefen, Rigdon and Straub [7] suggested that PLS and CB-SEM are “very different in their underlying philosophy, distributional assumptions and estimation objectives.” They go on to say that “PLS shines forth in exploratory research”, and in “situations that are 'data-rich but theory-primitive’”. Joreskog and Wold ([10] tell us that “ML [CB-SEM] is theory-oriented, and emphasizes the transition from exploratory to confirmatory analysis. PLS is primarily intended for causal-predictive analysis in situations of high complexity but low theoretical information.”

It is interesting to explore what these statements mean in the context of scientific research in general. It is possible to think of a continuum of theory development and testing that moves gradually from (1) pure observation with no theoretical preconceptions, to (2) the beginnings of some conceptual models and some targeted data collection where we look at data and apparent relationships to understand what is going on, revising our models as we go, to (3) where we have decided upon detailed hypothesized models and wish to test them statistically to see if they can stand up to patterns found in data from the real world.

Relative to the statements earlier about where PLS is strongest, two questions must be asked. First, what is the role of statistics and tests of statistical significance at the various stages of research? Secondly, where should CB-SEM or regression be used in these stages, and where should PLS be used?

It seems fairly clear from the requirements of statistical testing that “fishing expeditions” [where many constructs are measured and then apparent relationships are examined to see which constructs seem to be related] are not the appropriate activity for statistical testing. For example, one could take a dataset of measures of twenty possible causes of a phenomenon, find one of the twenty that seems to be significantly related at $p < .05$, and then statistically test that single cause as a possible predictor of the phenomenon. If you use the same dataset you started with, you will probably find that the one earlier statistically significant cause is again a statistically significant ($p<.05$) predictor of the phenomenon.

However, that statistical significance truly tells the researcher nothing scientific. Using a $p$ cutoff of .05, we would expect about 5%, or one of the original 20 possible causes, to be statistically significant by random chance, even if none have a true relationship. That one seemingly related cause could easily be no more related in the total population than the other 19. Using the same data in a second analysis assures that you will get the same result, even when there is no true relationship between the constructs. What would be informative is to collect new data on the one (or the twenty) constructs, and then do a statistical test with the new data. If this second test showed the same link between constructs to be statistically significant, it would tell us something.

The point of the above is that tests of statistical significance only have meaning when we have concrete hypotheses up front, and test these hypotheses with data that was not used in their development. This is fairly basic to the assumptions
underlying tests of statistical significance. In other words, producing a t statistic and a value for statistical significance is only appropriate in the third stage of theory development. This is not to say you can’t use a “fishing expedition” as a way to see what is happening in a domain, and even to look at the strength of apparent relationships based on t statistics. But unless there is new data (not already used to discover apparent relationships) to test hypothesized relationships, researchers should never present t statistics alone as evidence of a true link.

With this in mind, we need to consider what is being said when we assert that PLS “shines forth in exploratory research”. Wold [21], pp. 28-29) talks about gains PLS is able to provide by sacrificing optimum estimates in favor of “deliberate approximation” with few constraints. Two of these are quite relevant to our topic. (1) instant estimation: PLS is easy to set up for an analysis, and the computer analysis “is usually performed in a fraction of a minute of computer time.” This is admittedly an advantage when we are engaged in exploratory analysis and wish to try out many such models to see which fits the data best. On the other hand, the speed and power of computers today means that CB-SEM can also be performed about as rapidly, if that is a consideration.

The second advantage Wold mentioned is (2) dialogue with the computer: “Soft modeling is an evolutionary process. At the outset the arrow scheme is more or less tentative. The investigator uses the first design to squeeze fresh knowledge from the data, thereby putting flesh on the bones of the model, getting indications for modifications and improvements, and gradually improving the design. For example, the case values of an LV may show high correlation with an observable that hasn’t as yet been included among its indicators. Or the residuals of a causal predictive relationship may be highly correlated with an observable that could be included in an inner relation as an LV with a single indicator.”

There is no other way to characterize what Wold describes above as other than a “fishing expedition”. There is nothing wrong with fishing expeditions. Fishing expeditions may produce lots of interesting hypotheses. But when we publish a hypothesis and a statistical test of it in our major journals, we are hopefully not engaged in a fishing expedition (exploratory work). When we publish results with statistical significance, they should be the result of a confirmatory analysis, not the result of a “dialogue with the computer” as described by Wold above. Wold describes PLS as an excellent tool for exploration in realms where there is a lot of data but no clear idea of the relationships (perhaps with secondary data). One implication might be that one should use PLS for exploration, and CB-SEM or regression and an unanalyzed dataset for confirmatory hypothesis testing.

6. PLS is consistent in the sense of consistency at large

Consistency. In the quote in Section 3 above, Wold says that PLS rests content with consistency, albeit in the qualified sense of consistency at large. We need to consider what these two terms (“consistency” and “consistency at large”) actually mean. Consistency has a specific definition in statistics. From Pindyck and Rubenfeld [15, p. 30]: “β is a consistent estimator of β if the probability limit of β is β. [Italics in the original.]” And, same page: “Strictly speaking β converges to β if for any δ>0 the following is true:

\[
\lim_{N \to \infty} \text{Prob}(|\beta - \beta^*| < \delta) = 1
\]

Here, N refers to sample size. The above says that if you can get the difference between the estimate and the true value to be as small as you like, solely by increasing N sufficiently, then the estimate is consistent. This is mathematically true for both regression and CB-SEM, assuming the assumptions are met.

We have already noted that when there is measurement error, the stated regression assumptions are not met, and under these circumstances regression estimates have a bias, and therefore are not “consistent”, as shown by Nunnally and Bernstein [14]. As stated earlier, this bias is equal to the square root of the product of the X and Y reliabilities. If X and Y have the same reliabilities, the bias is equal to \(\alpha\), Cronbach’s alpha. Cronbach’s equation for reliability is:

\[
\alpha = \frac{K \cdot S}{1 + S (K - 1)}
\]

Note that K here is the number of indicators (as opposed to sample size in the earlier mentioned definition of consistency) and S is the average correlation between indicators.

As shown in this equation, the bias of regression does not decrease as sample size increases – there is no N in the equation. If there is bias due to measurement error, and that bias is not reduced by increasing sample size, then by the statistical definition described earlier, regression is not “consistent”.

4743
Neither is PLS consistent. As shown by Chin [1, p. 330], the bias in the PLS estimates for a two block model (two constructs, one predicting the other with both constructs having the same reliability) is equal to:

$$\text{PLS Bias} = \frac{S}{S + (1-S)/K}$$

Where S is the average correlation coefficient across all indicators, and K is the number of indicators. Note that in this equation, just as for regression, the bias does not decrease as sample size increases, because sample size is not part of the equation. Therefore Chin’s equation above is an indirect acknowledgement that PLS is also not “consistent” in the statistical sense described above.

We can go further and show that Chin’s equation for bias in PLS estimates has an interesting relationship to Nunnally and Bernstein’s [12, pp. 241, 257] equation for bias in regression. Assuming that X and Y both have the same reliability, (as we did earlier for regression) we can rewrite Chin’s equation as follows

$$\text{PLS Bias} = \frac{S}{S + (1-S)/K} \quad \text{Chin’s equation}$$

= $$\frac{K * S}{K * S + (1-S)/K}$$

= $$\frac{K * S}{[K * S + (1-S)]}$$

= $$\frac{K * S}{[K * S + 1-S]}$$

= $$\frac{K * S}{[1 + (K * S) - S]}$$

= $$\frac{K * S}{[1 + S (K - 1)]} = \text{Cronbach’s Alpha}$$

So we see that Chin’s equation for the bias in PLS estimates is exactly the same as Nunnally and Bernstein’s equation for the bias in regression estimates. Neither PLS nor regression is “consistent” when there is measurement error.

**Consistency At Large.** Returning to the question of what is meant by consistency at large, Joreskog and Wold state [10, p. 266, italics in the original]: “Under general conditions of regularity, ML estimates of the unknowns are known to be optimal in the large sample sense (asymptotic minimum of the confidence intervals), while the accuracy of the PLS estimates is less sharp – they are asymptotically correct in the joint sense of consistency (large number of cases) and consistency at large (large number of indicators for each latent variable).”

To test their assertion that PLS has consistency at large, Hui and Wold [9] carried out a Monte Carlo simulation based on the model shown in Figure 1 below.

**Figure 1. Replication of Hui and Wold’s Simulation, With Weights of .7 on X and .8 on Y**

Below we show results of a replication of that simulation, using 4, 8, 16 and 32 indicators as Hui and Wold did, and using 4000 cases. We have also included regression in the simulation and the figure.

**Figure 2. Consistency At Large**

**Hui and Wold’s (1992) Simulation Results**

(4, 8, 16, 32 indicators, N=4000)

Clearly the PLS and the regression lines are right on top of each other in Figure 2. Though neither PLS nor regression has reached the actual path value of .5, it is quite reasonable to assume that as the number of indicators increased further to 64 or 128 etc., the two lines would converge with the true value line. This is indeed evidence that PLS and regression have “consistency at large”.

Exploring a bit further, we note that at each of the four numbers of indicators, the difference between the estimate (both PLS and regression) and the correct value is exactly equal to that predicted by Nunnally and Bernstein’s [14] equation for measurement bias in regression, restated here:

$$\text{Reg Bias} = \beta_{XY}^* / \beta_{XY} = \text{SqrT} (\alpha_X \cdot \alpha_Y)$$
This suggests that the only bias when very large samples sizes are used (for both PLS and regression) is due to measurement error.

Decomposing Consistency At Large. It is interesting that for PLS to be consistent at large, two different conditions are required. It is illuminating to consider what is accomplished by each of these conditions separately. First consider what “large number of indicators for each latent variable” actually means. A quick look at Cronbach’s equation for alpha shows that as the number of indicators per construct (K in Cronbach’s equation) approaches infinity, \( \alpha \) approaches one. In other words, increasing the number of indicators reduces measurement error, and increasing the number of indicators to infinity reduces measurement error to zero.

Where Figure 2 allowed us to see the effect of increasing the number of indicators holding sample size constant at 4000, Figure 3 allows us to see the impact of having a “large number of cases”, versus a small number. More specifically, it shows the effect of moving from a sample size of \( N = 4000 \) down to \( N = 100 \). The results are interesting.

| Figure 3. PLS & Regression, Moving from \( N = 4000 \) to \( N = 100 \) |

What we see in Figure 3 is that as the sample size decreases from 4000 to 100, there is a small change in the regression estimates at 16 or 32 indicators, but PLS has a considerably larger change (an increase) across each condition of number of indicators. What explains this increase? It is hard to argue that the PLS estimates with \( N = 100 \) are more accurate than with \( N = 4000 \), since consistency at large has already said that with much higher numbers of cases, PLS will be closer to the true value. Furthermore, at \( N = 100 \) and 32 indicators, PLS clearly overestimates the true value.

Making Sense of the Impact of Many Cases. Following Rönkkö and Ylitalo [18], we can think of the PLS estimation process as proceeding in three distinct steps. First, the weights and loadings of the indicators are determined in a unique PLS algorithm, and construct scores are determined from those. Secondly, those constructs scores are used in a traditional ordinary least squares regression to determine path values between constructs. Thirdly, bootstrapping is used to determine the standard deviations of the path values, and the t statistics.

Thus the key difference between regression and PLS is the way the weights and loadings are determined in step one. If PLS path estimates are at all different from regression in a substantive way, it is because PLS comes up with different weights and loadings for use in determining construct values. Otherwise PLS and regression would always find the same path estimates.

This raises the possibility that PLS might be capitalizing on chance in the following way. When there is measurement error, each indicator has random variance around the true score of its construct, and random variance around the correlations with other indicators. Some indicators of the X construct will be (by random chance) more correlated with the Y construct than others. Likewise, (again by random chance) some indicators of Y will be more correlated with the X construct than others. If PLS were to weight those highly correlated indicators more heavily, it would tend to overestimate the relationship between X and Y.

Presumably, the more random variance in the picture, the more variety PLS would have to choose from, and the more extensive the opportunity for overestimation would likely be. Increasing the number of cases would certainly decrease the amount of random error in the indicator correlations that would be available to PLS as it determined the weights for each construct’s indicators.

One possible explanation of the increase of the PLS estimates shown in Figure 4 is that with smaller sample size, there is more random variance in the estimates of the indicator correlations, and more opportunity for PLS to capitalize on chance. This is a fairly startling possibility. However, we are unable to offer any other persuasive explanation.

Figure 4 below represents our interpretation of the findings of our Monte Carlo simulation. Both PLS and regression suffer equally from the bias caused by measurement error. The more indicators, the higher the reliability, and the smaller the bias due to measurement error. However with a sample size of 100, PLS also seems to take advantage of random variations in indicator correlations, and capitalizes on...
chance by giving higher weights to some indicators to increase the $R^2$ values of its constructs. This increases the PLS path estimates (for any number of indicators), even increasing it so that it overestimates the true value when there are very many indicators. With very large sample sizes, random variations in the indicator correlations are reduced considerably, and PLS does not capitalize on chance, giving exactly the same path estimates as regression.

Figure 4. When Sample Size Decreases From 4000 to 100, PLS and Regression Suffer Equally From Measurement Error. PLS Also Suffers From Capitalization On Chance

7. Discussion and Conclusions

We began this paper by noting that over the past 15 or 20 years, PLS has become the statistical technique of choice in the MIS research community [17]. Though the original developers and users of PLS described it as being appropriate for exploratory work, discovering potential relationships and emphasizing prediction over explanation [21], over time MIS researchers have increasingly adopted the technique for confirmatory (theory testing) contexts.

There have recently been a number of challenges to the efficacy of PLS. Studies have raised questions about perceived advantages of PLS under conditions such as small sample sizes [8], the use of formative measurement [14] and the presence of measurement error [18]. Relevant to the issue of theory testing, Evermann and Tate [5] demonstrated that the typical criteria used for evaluating the “fit” of a PLS model to the data are generally ineffective, and concluded that “…the use of the PLS path modeling as a tool for statistical hypothesis testing should perhaps be reconsidered, although more analysis on the PLS method is clearly in order.”

In this paper we have focused on four additional considerations with respect to the use of PLS. We argue that (1) rather than having more statistical efficacy than other techniques when data doesn’t conform to the certain distributional requirements, PLS in fact has less statistical efficacy when data does conform to assumptions, and that nothing can really be said mathematically about PLS’ characteristics when data does not conform to assumptions. This means that PLS sacrifices the possibility of optimal solutions, which are generally characteristics of widely used estimators in scientific research (even if some assumptions are not always met). (2) If a researcher reports a statistical test of a hypothesis with PLS, they are engaging in confirmatory (not exploratory) work. This is not the domain where PLS is reputed to have its greatest strengths. (3) The pseudo F test for comparing PLS models is at best an approximation that will be biased by changes in the indicator weights and loadings.

Finally, (4) by decomposing the “consistency at large” conditions, we have been able to isolate the impact of measurement error on one hand, and perhaps also isolate the impact of capitalization on chance on the other hand. This raises the disturbing possibility that the biggest difference between PLS and regression is that at sample sizes typically used in MIS research, PLS is capitalizing on chance. (Our simulation looked specifically at sample sizes of 100.) Obviously there needs to be more work done to confirm this finding, and to determine under what circumstances it is aggravated.

What are the practical implications for MIS researchers contemplating analyzing a path model? Among the authors of this paper, we see two possible perspectives on this. One perspective is that repeated empirical tests have demonstrated that under many of the conditions typically experienced by MIS researchers, PLS path estimates do seem to come very close to regression estimates (though in comparison to CB-SEM both seem to ignore measurement error in their path estimates). PLS also typically displays similar power [8] to regression and CB-SEM. From a pragmatic standpoint, PLS is easier to use than regression or CB-SEM techniques in situations where there are large, complex models (when there are numerous indicators and constructs with multiple mediating relationships, etc.). If this is true, and if the conditions under which PLS capitalizes on chance enough to seriously overestimate path values are quite unusual in practice (for example with very low indicator loadings and small path effect sizes), perhaps researchers would be willing to accept the trade-offs of more ease of use even if it came with the potential for slightly biased estimates. This is especially true when we consider that most MIS research is focused on determining
whether or not a path value is greater than zero in a statistically significant sense, rather than on obtaining accurate path values.

A different perspective is the following. PLS was introduced to and advocated to MIS researchers on the basis of a list of supposed strengths that were very attractive: better at small sample size, better with non-normal distributions, better with formative constructs, ability to easily and correctly take into account measurement error in one’s estimates. Empirical work has now stripped away most of the dreams about PLS being a silver bullet that improves on other existing statistical techniques. Though more work is needed to confirm this, we also believe PLS does capitalize on chance in a way not seen in regression or CBSEM, and that this explains the much of the difference between PLS and regression path estimate values.

The question one could ask is this. If MIS researchers could go back in time, knowing what we know now, would we turn away from tried and true techniques like regression and CB-SEM (that do have optimal qualities) and begin to use a technique that might legitimately be called “pseudo” regression? If we did this, what would we expect scientists in other disciplines to think about our seriousness as researchers? Would continuing to endorse PLS have negative impacts on the MIS field’s reputation as a field of serious academic research?

Regardless of the perspective taken by MIS researchers going forward, we believe there is sufficient empirical evidence to question the seemingly enthusiastic embrace of PLS path modeling for confirmatory analyses by MIS researchers. As such, we are recommending that PLS be considered a complementary technique (rather than the primary choice), and that there be a call for more careful analysis of PLS under the conditions typically seen in MIS research.

9. References


