Communications of the Association for Information Systems

1-2015

Sample Size Determination and Statistical Power Analysis in PLS Using R: An Annotated Tutorial

Miguel Aguirre-Urreta

Information Systems and Quantitative Sciences, Rawls College of Business Administration, Texas Tech University, USA,
miguel.aguirre-urreta@ttu.edu

Mikko Rönkkö

Institute of Strategy and Venturing, Department of Industrial Management, School of Science, Aalto University, Finland

Follow this and additional works at: https://aisel.aisnet.org/cais

Sample Size Determination and Statistical Power Analysis in PLS Using R: An Annotated Tutorial

Miguel I. Aguirre-Urreta

Information Systems and Quantitative Sciences, Rawls College of Business Administration, Texas Tech University, USA

miguel.aguirre-urreta@ttu.edu

Mikko Rönkkö

Institute of Strategy and Venturing, Department of Industrial Management, School of Science, Aalto University, Finland

Abstract:

Partial least squares (PLS) is one of the most popular analytical techniques employed in the information systems field. In recent years, researchers have begun to revisit commonly used rules-of-thumb about the minimum sample sizes required to obtain reliable estimates for the parameters of interest in structural research models. Of particular importance in this regard is the a priori assessment of statistical power, which provides valuable information to be used in the design and planning of research studies. Though the importance of conducting such analyses has been recognized for quite some time, a review of the empirical research employing PLS indicates that they are not regularly conducted or reported. One likely reason is the lack of software support for these analyses in popular PLS packages. In this tutorial, we address this issue by providing, in tutorial form, the steps and code necessary to easily conduct such analyses. We also provide guidance on the reporting results.

Keywords: Partial Least Squares, Statistical Power, Monte Carlo Simulation, Sample Size, Research Planning.

Sample Size Determination and Statistical Power Analysis in PLS Using R: An Annotated Tutorial

I. INTRODUCTION

For a quantitative researcher, being able to provide a solid statistical foundation for research findings is of central importance. Among many issues related to research planning and design, statistical power considerations, which influence the minimum sample size that should be obtained to provide a desired likelihood of finding an effect when that effect indeed exists in the population—a central goal of empirical research—are an important, if somewhat overlooked, issue (Baroudi & Orlikowski, 1989). In particular, analyses based on partial least squares (PLS) are very common in the field. Given the commonly held belief that PLS is an appropriate statistical choice for small samples, and that sample size is a central factor in the determination of statistical power, the conduct of a priori power analyses when PLS will be employed would seem to be particularly relevant to planning research. Though the need for this has been evident for some time with repeated calls to attention to this issue (Marcoulides, Chin, & Saunders, 2009; Marcoulides & Saunders, 2006), it is the case that few studies employing PLS report—or conduct—power calculations, even though methodologists strongly recommend them (Baroudi & Orlikowski, 1989; Chin, 1998; Marcoulides et al., 2009; Marcoulides & Saunders, 2006). As Kaplan (1995, p. 117) notes: "Assessing power, therefore, must now become a routine part of establishing the statistical validity of an estimated model".

In order to ascertain the degree to which statistical power issues are regularly considered in research employing PLS as the statistical technique of choice, we reviewed all studies identified by Ringle, Sarstedt, and Straub (2012), which cover empirical research using PLS published in the 1992-2011 period in *MIS Quarterly*. Out of the 65 studies identified therein, only a small proportion (11, or 17 percent) contained any mention of statistical power. Moreover—as we discuss later in Section 3—all of these studies present some shortcomings in the way in which authors performed and/or reported power calculations. Therefore, we conclude that the state of practice in power assessment in studies employing PLS could be improved.

With this tutorial, we provide guidance on why considering statistical power is important, and on how statistical power can be assessed when using PLS. We begin by briefly introducing the statistical power concept in Section 2, in which we place particular emphasis on the negative consequences of both underpowered and overpowered research designs. Next, in Section 3, we review the different approaches to power calculations that have been so far suggested or employed—even if in a limited capacity—in studies using PLS, and note limitations present in each one of them. We then introduce in Section 4, in tutorial form, the calculation of statistical power using the open source statistical environment R, and provide researchers with sample code for conducting a power analysis. In Section 5, we conclude with recommendations for the conduct and report of power analyses in studies employing PLS.

II. STATISTICAL POWER AND RESEARCH PLANNING

Statistical power is associated with frequentist inference, most typically null-hypothesis significance testing (NHST). In NHST, researchers obtain a measure of the probability of obtaining a particular estimate, or a more extreme one, if the null hypothesis of no effect is true in the population, which is given by the p value used in statistical analyses. Statistical power refers to the probability of correctly rejecting a null hypothesis when that hypothesis is not true in the population. A related concept is significance level, which is the probability of rejecting a null hypothesis when that hypothesis is indeed true in the population. In both these cases, probability is defined as the limit of relative frequency in a large number of replications.

These two concepts are closely linked to type I and type II statistical errors, which refer to false positive and false negative findings, respectively (see Table 1 for summary). Consider, for example, two variables: A and B. If these two variables were correlated in the population of interest, but we were to conclude that our analysis does not provide statistically significant support for the correlation, we would be committing a type II error. The likelihood of this kind of error is directly dependent on the power of our research design and analysis.

Table 1: Type I and Type II Errors			
		Population	
		Null hypothesis is true	Alternative hypothesis is true
Sample	Accept null hypothesis	Correct decision	False negative Type II error
	Reject null hypothesis	False positive Type I error	Correct decision

Most, if not all, IS studies using PLS employ the technique on survey data to test if a hypothesized relationship is non-zero in a population. This kind of research design has two important parameters that affect statistical power, and which are under the control of researchers: sample size, and which and how many indicators are used in the study. Generally, a larger sample size increases the accuracy of a single point-estimate and so decreases its standard error. When the estimates are more accurate and there is less uncertainty about their accuracy (i.e., smaller standard errors), we can be more confident that, if our estimate is positive, the population value is also positive. The number of indicators and their quality affects power, too: in structural equation models, the accuracy of the estimates of the relationships between the latent variables depends not only on the sample size, but also on the quality of the measurement model. Increasing the number of indicators and their quality will also increase the accuracy of the estimates and decrease their standard errors.

These are not merely theoretical considerations, but have practical relevance for researchers. Consider, for example, a case where a researcher conducts a power analysis and concludes that the study design has only .50 power to detect an underlying relationship given the assumptions that the researcher has made about the effect in the population. In this case, the researcher faces a choice: 1) to either pursue the study as planned knowing that there is a 50% chance to not find an effect even if the theory was correct, and consequently likely not being able to publish the study, or 2) to plan for more and better indicators and a larger sample size to increase power and make the odds more favorable.

Problems with Underpowered Studies

The most evident issue with underpowered studies is their inability to consistently detect the presence of an effect when that effect actually exists in the population. In addition to the deleterious impact on cumulative research and knowledge advancement, this issue also hinders the possibility that the findings will ever be published, given the general bias against publishing non-significant results in the organizational and social sciences. On this count alone, it is important for researchers to ensure that their studies have sufficient statistical power to detect the effects of interest.

There is a second problem with underpowered studies that, however more nuanced, is not less important. As it is well known, the smaller the sample size used to obtain an estimate, the larger the variability in the estimates being obtained for a parameter of interest. Consider, for example, the case of the correlation between two variables. In this example, assume that the true population value of the correlation is 0.30. A researcher who collects data on these variables will be interested in comparing the sample estimate of the correlation against the null hypothesis that the two variables are uncorrelated. In other words, a researcher would be interested in assessing the probability of observing an equally high correlation just by chance. To illustrate this, we have plotted the sampling distribution of the correlation of two uncorrelated random variables under three different sample sizes in Figure 1 and shaded the area in which the sample correlation falls 95% of the time.
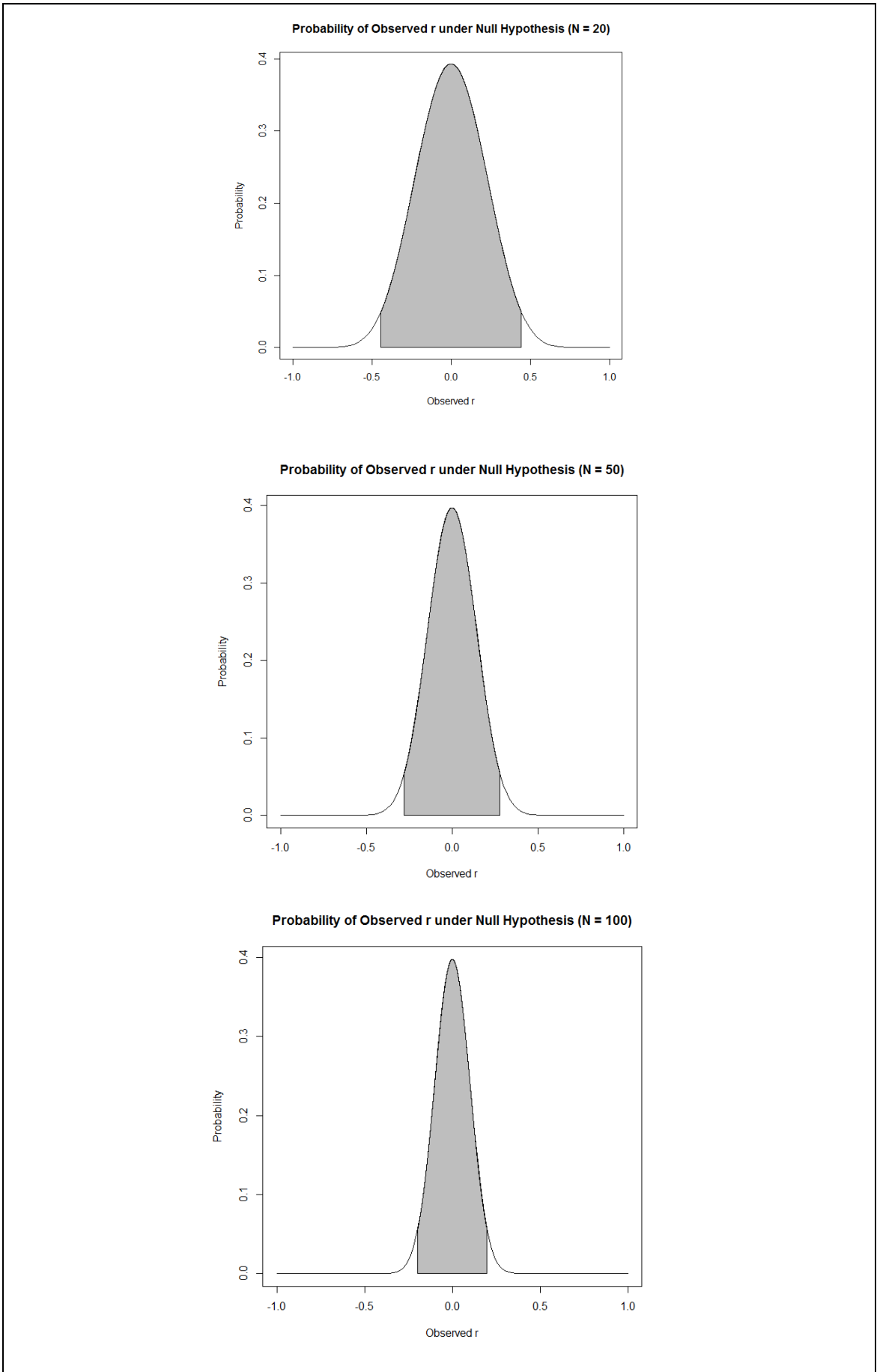
Figure 1. Probability Distributions for Observed r

The figure shows that, with the smallest sample size, even fairly large sample correlations are possible even if the variables were unrelated in the population. The exact distribution function of correlation is complex and, therefore, instead of assessing the magnitude of the correlation directly, we use the relationship of the correlation to its standard error as a test statistic. This statistic is known to be distributed as a *t* with degrees of freedom equal to N − 2 (Cohen, Cohen, West, & Aiken, 2003, p. 49):

$$t = \frac{r}{\sqrt{(1 - r^2)/(N - 2)}}$$

With this *t*-statistic we can calculate exact values that the correlation must exceed to conclude that there is less than 5 percent chance to obtain a similarly high value just by chance. With a sample size of twenty, rejecting the null hypothesis at a 5 percent level of significance would only occur if the sample estimate of the correlation was 0.4438 or larger; that is, at least 47.9 percent larger than the true population value of the statistic. For a sample of fifty, the critical value is 0.2787, and, for a sample of one hundred, 0.1984.

The point of the example is that small samples require much larger critical values before the estimates are considered significantly different from zero and, thus, rejecting the null hypothesis of no correlation. This, coupled with the existence of a publication bias toward significant results—also known as the "file drawer problem" generally discussed in the literature on meta-analysis—indicates that underpowered studies that do get published are likely overstating the true effect in the population. This is an issue that has been discussed in the medical literature, where the overstatement of a treatment effect has major practical implications (Ioannidis, 2008; Yarkoni, 2009), but it has also started to make its way into the social and organizational sciences (e.g., Hollenbeck, DeRue, & Mannor, 2006). The issue is even more complex for multiple regression models because the additional complexity involving multiple predictors is likely to result when small samples are used in patterns of findings that are very idiosyncratic to the particular sample under analysis (Maxwell, 2004). Given the prevalent belief in the IS field that PLS is best suited for analyses with small samples, these issues should be of major concern to researchers in the field.

Problems with Overpowered Studies

The most evident problem arising from the conduct of overpowered studies is the unnecessary consumption of the time and resources involved in obtaining an overly large sample. While the issue may not be major when employing readily available populations, such as university students, for conducting research, it takes on more importance when other groups are targeted. At the extreme, if the conduct of the research involves some discomfort to the respondents (either physical or psychological), the ethics of subjecting more individuals than strictly necessary to the research must be considered. What degree of statistical power—and hence sample size—should be deemed acceptable requires the balancing the practical implications of committing type I and type II errors and what do those imply for the findings of the research. The commonly used guideline in the organizational and social sciences of a significance level of five percent and statistical power of eighty percent should therefore be considered on a case-by-case basis. However, absent a priori power calculations, researchers may be collecting much more data than is necessary to conduct their study when very large samples are obtained.

III. STATISTICAL POWER AND PLS

Estimating the power level of a study is a key characteristic of rigorously applying SEM (Gefen, Rigdon, & Straub, 2011), and at least one recent paper giving guidelines on using PLS (Peng & Lai, 2012) explicitly states that a power analysis should be performed. Unfortunately, in contrast to many CBSEM packages that include Monte Carlo simulation features that can be used for power analyses, none of the widely used PLS applications provide anything similar. Due to lack of software support, most assessments of power are based on rules of thumb that have not been validated or on applying power analysis procedures that were designed for multiple regression and thus assume that all variables are measured without error. We now return to the 11 identified studies from the review by Ringle et al. (2012) that do discuss statistical power to some degree, and explain how these studies addressed the issue, and why these power analyses were either inappropriate or incorrectly implemented. These studies can be divided into three categories: those arguing for sufficient power based on sample size heuristics, those using analyses and recommendations for regression analysis, and those using Monte Carlo simulations.

Many of these studies discuss power, but do not report how power was assessed. Enns, Huff, and Higgins (2003), when discussing a non-significant finding, note that power analyses suggested the sample size was sufficient. As before, however, they provide no details about how they made this determination. Lewis, Agarwal, and Sambamurthy (2003), on the other hand, caution readers when interpreting their non-significant findings because their sample size was only a fraction of that which, by their estimate, would be necessary for adequate power, though they provide no supporting details about power calculations. Venkatesh and Morris (2000) cite the high power of a test as additional support for their hypothesis of no significant effects; however, how they do not discuss

how they carried out those power calculations in any detail. Particularly problematic is a statement from Wells, Valacich, and Hess (2011, p. 388): "In that all effects were significant at the alpha protection level of .05, statistical power and Type II errors are not issues". Rather, as we discuss earlier, significant results arising from underpowered studies may present estimates of effects that are severely biased when compared to their population counterparts. In fact, statistical power is an important factor in conducting research, regardless of whether the results turn out to be significant when data are collected because power calculations should be performed beforehand. Work by Jiang and Benbasat (2007) represents an exception in that their power analysis does not have any obvious issues. However, they only consider statistical power for the between-group comparisons in their research, conducted with ANOVA procedures, but not for their PLS analyses.

Determination of Power Based on Sample Size

Two studies use sample size as a justification for sufficient power, but neither provides a clear reference to any guideline on sample size. Ang and Straub (1998) simply state that a sample size of at least two hundred would be enough to ensure sufficient statistical power, and that the sample actually obtained—two hundred and twenty five cases—indicated high statistical power. He does not provide details about how he determined this sample size, what effect sizes he evaluated, and which particular effect he analyzed. Similarly, Siponen and Vance (2010) use sample size as a justification without reporting any analyses to support their conclusions. They justify their approach by citing Marcoulides and Saunders (2006) that nevertheless does not provide any explicit sample size recommendations, but rather suggests that Monte Carlo simulations should be used to estimate power.

In general, using any rule-of-thumb for sample size planning or assessing statistical power is problematic. Consider the commonly used guideline the "rule of ten" introduced by Chin and Newsted (1999). In the commonly accepted interpretation of this guideline, researchers should consider either the composite with the largest number of formative indicators or the composite with the largest number of incoming paths from other composites in the model, and multiply the largest number of these paths by ten. In the past, researchers have used this rule to justify the estimation of models with very small samples, which would not otherwise be considered acceptable in terms of statistical power and stability of the estimates (e.g., magnitude of the resulting standard errors). This state of affairs led to the editorial comments by Marcoulides and Saunders (2006), who admonish researchers to "Please make sure that your sample size is large enough to support your conclusions and do not allow perceived fancy modeling techniques to overwhelm common sense" (p. viii). More recently Marcoulides, Chin, and Saunders (2009, p. 174) explicitly state that the "rule of ten" was not meant to be a guideline that should be followed: they emphasize that Chin and Newsted (1999) actually conclude that, "for a more accurate assessment, one needs to specify the effect size for each regression analysis and look up the power tables" (p. 327). However one wishes to interpret the meaning behind these statements, it is clear that the "rule of ten" has no basis on statistical estimation, nor is a sound basis for assessing statistical power. Furthermore, it does not take into account other characteristics of the research design, such as the reliability of the items attached to each involved composite or their number, which also affect the likelihood of detecting an effect. In this respect, we support the repeated calls to abandon the use of this rule-of-thumb unconditionally across all situations (Marcoulides et al., 2009; Marcoulides & Saunders, 2006). Gefen, Rigdon, and Straub (2011, pp. A3-A4), however, seem to cautiously recommend the "rule of ten", although with the caveat that it is not supported by any research.

Power Analysis Using Recommendations or Methods for Regression Analysis

Power analysis using guidelines and methods developed for regression analysis was the second most common option among the 11 studies. Klein and Rai (2009) assessed the power of their statistical tests by following the procedures outlined by Cohen (1988) for multiple regression analyses, but it appears the authors assessed power using the actual estimates obtained (e.g., a post hoc or retrospective power calculation), which is a rather problematic practice (Zumbo & Hubley, 1998). Liang, Saraf, Hu, and Xue (2007) also employ power tables by Cohen (1988) on the assumption that those would be applicable to PLS because PLS is a form of multiple regression. However, as we note later, the power tables do not take into account many of the specific characteristics of the PLS algorithm, nor the appropriate level at which statistical inferences are made by researchers. The only study that used a specialized power analysis software was that by Furneaux and Wade (2011), in which the authors used G*Power. However, the authors only consider the power of their test for the presence of common method variance, but not for the main model of interest.

The most common of these approaches adopted from the literature on regression analysis is to use power tables for multiple regression such as those provided by Cohen (1988) in his seminal work on power analysis. In this approach, researchers should ascertain the portion of the PLS model with the most regression paths, assume an effect size of interest (using, for example, Cohen's f^2 and associated guidelines as to what magnitudes of the statistic can be considered small, medium, or large effects), then look up the appropriate sample size for the desired power level in the tables provided. Though this approach is certainly an improvement over the rule of thumb we

previously discuss, it nonetheless suffers from two important weaknesses that render it less accurate than when first considered.

First, the commonly used measure of effect size f2 is a function of the variance explained in the dependent variable of interest (Cohen, 1992); that is (where values of 0.02, 0.15, and 0.35 are generally considered to be small, medium, and large effects):

$$f^2 = \frac{R^2}{1 - R^2}.$$

The proportion of variance explained in a variable by its predictors in a regression model is a function of the strength of the correlation between the predictors and the dependent variable and the strength of the correlation of the predictors among themselves. The relationship between the number of predictors and the magnitude of the overall effect so defined is, however, only an approximation because the power tables make no allowance for predictors of different magnitudes. Consider the following example: a regression with three predictors, assuming a large effect size and desired power of 0.80, would require a sample size of thirty four (Cohen, 1992, Table 2). Note that a large effect size (e.g., an f2 of 0.35) is tantamount to 25.93 percent of explained variance in the dependent variable. For simplicity, assume the predictors are uncorrelated among themselves. In this scenario, the contribution of each predictor to the explained variance is the square of its correlation with the dependent variable.

Therefore, an explained variance of 25.93 percent can be obtained with one predictor correlated at 0.40 with the dependent variable and two predictors correlated at 0.223, or with three predictors that correlate at 0.294 with the dependent variable, or many other combinations. In the first case, the sample size that corresponds to a given level of statistical power to detect a correlation of 0.40 (for the first predictor) cannot logically be the same that is needed to detect a correlation of 0.223 (for the other two predictors). Though Cohen (1988) notes that the unique variance explained by any single predictor should be considered when estimating statistical power, it is nonetheless the case that the power tables provided only consider two inputs when establishing the sample size necessary to achieve a particular level of statistical power: the number of predictors, and the magnitude of the overall effect size. Moreover, when it comes to significance of estimates, researchers are typically more concerned with the statistical significance of each individual predictor than with that of the explained variance overall. In fact, we are not aware of any study in the field that has examined the significance of the overall variance explained in a dependent variable. Therefore, sample size calculations that focus only on this statistic (the test statistic considered in the power tables is the overall F test for the entire regression model, which is typically not interpreted in research studies) and not on the individual magnitude of the relationships between the predictors and the dependent variable—which is at the level which statistical significance will later be assessed—can only be considered approximations.

The second issue with using tables from multiple regression models or any other tool that is designed for assessing power of multiple regression is that they assume perfect reliability in the variables involved, whereas it is the case in PLS that the composites used to represent the constructs of interest, being themselves weighted sums of indicators containing measurement error to varying degrees, are not perfectly reliable. Because of this, structural estimates from PLS analyses are biased to some degree when compared to their population counterparts, which further complicates sample size calculations. In addition, because the PLS algorithm takes advantage of sampling variability in the iterative calculation of the weights assigned to each indicator (Aguirre-Urreta, Marakas, & Ellis, 2013; Rönkkö & Evermann, 2013; Rönkkö, 2014), and this affects the obtained results, any sample size calculations for PLS need to take into account these idiosyncrasies of the algorithm, and power tables constructed for multiple regression models with perfect reliability do not fully capture them. Both these factors lead to considerably overestimating statistical power, which we show in Section 4.

Power Analysis with Monte Carlo Simulations

A third alternative, and the one on which we focus for the rest of this paper, is conducting proactive Monte Carlo studies as a way of determining the appropriate sample size given a target population model and various assumptions as to the distribution of the variables and the balance between type I and type II errors (Wolf, Harrington, Clark, & Miller, 2013). In this approach, which Marcoulides and Saunders (2006) also recommend, researchers first specify a target population model about which statistical inferences will be made, and then draw repeated data samples from that model—attending to any desired distributional characteristics of the involved variables—and analyze each of those samples with the desired analysis model and statistical technique. The proportion of times over all those replications that a particular parameter is found to be significant is an estimate of the statistical power of the sample size used to draw the replications. The simulation can be repeated with varying population values for the parameters of interest and different sample sizes to draw a picture of what to expect in terms of statistical power and estimate precision for a particular research scenario.

In their editorial comments, following earlier work by Muthén and Muthén (2002) on the use of Monte Carlo simulations for statistical power analyses and sample size determination, Marcoulides and Saunders (2006) introduce this approach as a valid alternative for determining an appropriate sample size, note the existence of several software tools that can be used to carry out such analyses, and present an example using Mplus (Muthén & Muthén, 1998). Most importantly, Marcoulides and Saunders (Marcoulides et al., 2009; Marcoulides & Saunders 2006) note that sample size determination for PLS analyses should consider various factors, such as distributional assumptions, psychometric characteristics of the indicators, or the magnitude of the relationships of interest, which paints a much more complex—but accurate—picture of the issue than previously used alternatives in this regard. However, Marcoulides and Saunders' (2006) recommendation is problematic in that they recommend using CBSEM analyses to estimate the statistical power for PLS. This approach is not entirely correct because the two families of techniques are different in how data are analyzed. In Section 4, we show that estimating the Monte Carlo samples with PLS and CBSEM will indeed lead to different results and conclusions.

Of the reviewed studies, only Majchrzak, Beath, Lim, and Chin (2005) performed simulations to determine statistical power, and they correctly used PLS to estimate the replicated model. Their simulation showed that their sample size of 17 provided sufficient statistical power. Considering that Cohen's (1992) power tables, which assume perfect measurement of the variables, state that the sample size required to obtain power of .80 for large effect size is always at least 30, stating that just over half of that would be sufficient under a scenario where measurement error is assumed to exist sounds unconvincing. Indeed, there seems to be an issue with their calculations: Majchrzak et al. (2005, p. 660) report that they "used bootstrapping with a sample size of 500". The idea of bootstrapping is to draw samples of the same size as the original sample, and, in this way, estimate the sampling variability of a statistic. If we increase the size of the bootstrap sample, the variance of the bootstrap estimates clearly decreases and the results are no longer a good approximation of how a statistic would behave when estimated from repeated samples of the original size. This is best explained with an example: consider a population with two uncorrelated and normally distributed variables. If we draw a large number of samples of 17 from this population and calculate a correlation between the two variables from each sample, the standard deviation of these estimates is about 0.25. If we further draw many bootstrap samples of 500 observations each from each of these samples of 17 and calculate the correlation from the bootstrap samples, the standard deviation of these correlations is, on average, about 0.04. Using this number as an estimate for the real sample variability of the correlation (which is 0.25) is clearly inappropriate, as would be using these in a significance test: if these bootstrap samples were used as part of a significance test, the analysis would result in a false positive rate of 75 percent, which might be mistakenly taken as a power estimate. (Appendix B provides R code for this short analysis). Further, the specific values assigned to each parameter in the simulation model by Majchrzak et al. (2005) were the actual estimates and not a priori specified ones, which makes this a post hoc power analysis, which is problematic in what can be inferred from the results (Zumbo & Hubley, 1998).

Monte Carlo simulations are currently the state-of-the-art in power analysis (Wolf et al., 2013). However, the current literature is problematic for PLS users. On the one hand, the literature suggests that CBSEM simulations can be used as approximations for the power of PLS. However, PLS is often used in cases where researchers argue that CBSEM would not be appropriate, and, if a researcher were to argue that CBSEM would not work well for a particular problem, it would be difficult to argue that it would work for a power analysis for the same problem. Moreover, some researchers (Reinartz, Haenlein, & Henseler, 2009) argue that PLS has more statistical power than CBSEM and, if this is indeed the case, then CBSEM power estimates would be biased approximations of the power of PLS. Finally, the idea that CBSEM simulations would be appropriate for estimating power of PLS analysis was introduced to the literature in an editorial that did not provide any supporting evidence, which makes it difficult to verify this claim. Then, on the other hand, the only study that correctly uses PLS to estimate the statistical power of PLS clearly uses some inappropriate statistics and makes a surprising claim that a sample as small as 17 provides sufficient power. This is problematic because it may lead researchers to question the usefulness of the Monte Carlo analysis altogether, while the real problem is in the inappropriate use of bootstrapping.

Summary of Current Practices

Taken together, our review indicates that statistical power considerations do not figure prominently in the literature employing PLS as the statistical technique of choice. Furthermore, in the handful of studies that discuss statistical power, the techniques employed were only either approximations to the procedure that we present next, or insufficient detail was provided to understand how the calculations themselves were performed, which effect sizes were considered, and which approach was employed. As we discuss in Section 4, the a priori calculation of statistical power for a variety of effect and sample sizes in a way that reflects the technique that will actually be used to analyze the data—PLS in this case—is an important element of research planning and design, and provides valuable information that helps assess the findings obtained later. We believe one important reason underlying this state of affairs is the lack of readily available procedures for calculating statistical power in a way that considers the characteristics of the PLS technique. Indeed, none of the commonly used PLS software packages (e.g., PLS-Graph

or SmartPLS) provide any support for power calculations, and researchers are left to their own devices about how to perform them. In this work, we remedy that situation by providing researchers with such an approach in a way that can be easily adapted to suit any research design.

IV. ANNOTATED TUTORIAL USING R

Introduction to R

We now illustrate how proactive Monte Carlo simulations can be used for power analysis with PLS. Our example uses matrixpls (Rönkkö, 2014), a new R package implementing the PLS algorithm, which is designed specifically for Monte Carlo studies using the powerful simsem (Pornprasertmanit, Miller, & Schoemann, 2013) package as a simulation framework. Following the example requires no prior knowledge of R and can be implemented in a "copy-paste" manner, although knowledge of R can be useful when seeking to modify and adapt the code included here for other research designs.

Step 1: Model Definition

The first step in a proactive Monte Carlo simulation is to completely specify the population model about which the researcher is interested in assessing statistical power. This requires that every parameter in the model be given a specific value. Sensitivity analyses can be conducted by making changes to the model and rerunning the simulation to see how those changes affect the results. In this tutorial, we use a four-construct model as our running example. In this example, we use the free Ωnyx software (von Oertzen, Brandmaier, & Tsang, 2015) as a graphical tool to construct our model. Although Ωnyx has fairly basic SEM estimation and simulation capabilities already built-in, its usefulness for our example comes from its ability to dynamically generate model syntax in OpenMX, Lavaan, Mplus, or RAM, or export the covariance matrix of all the observed and latent variables. Figure 2 shows our population model drawn in Ωnyx with the generated Lavaan model syntax on the right. Note that all relationships in the model have been assigned a specific value. When planning a new study, researchers should derive these based on past research, theoretical expectations as to relative magnitudes of effects, and experience with the research field. The particular constructs included in the model and the posited relationships among them should be derived from the theory and hypotheses underlying the research design. As we note above, several models can be specified with different values assigned for each of these parameters.
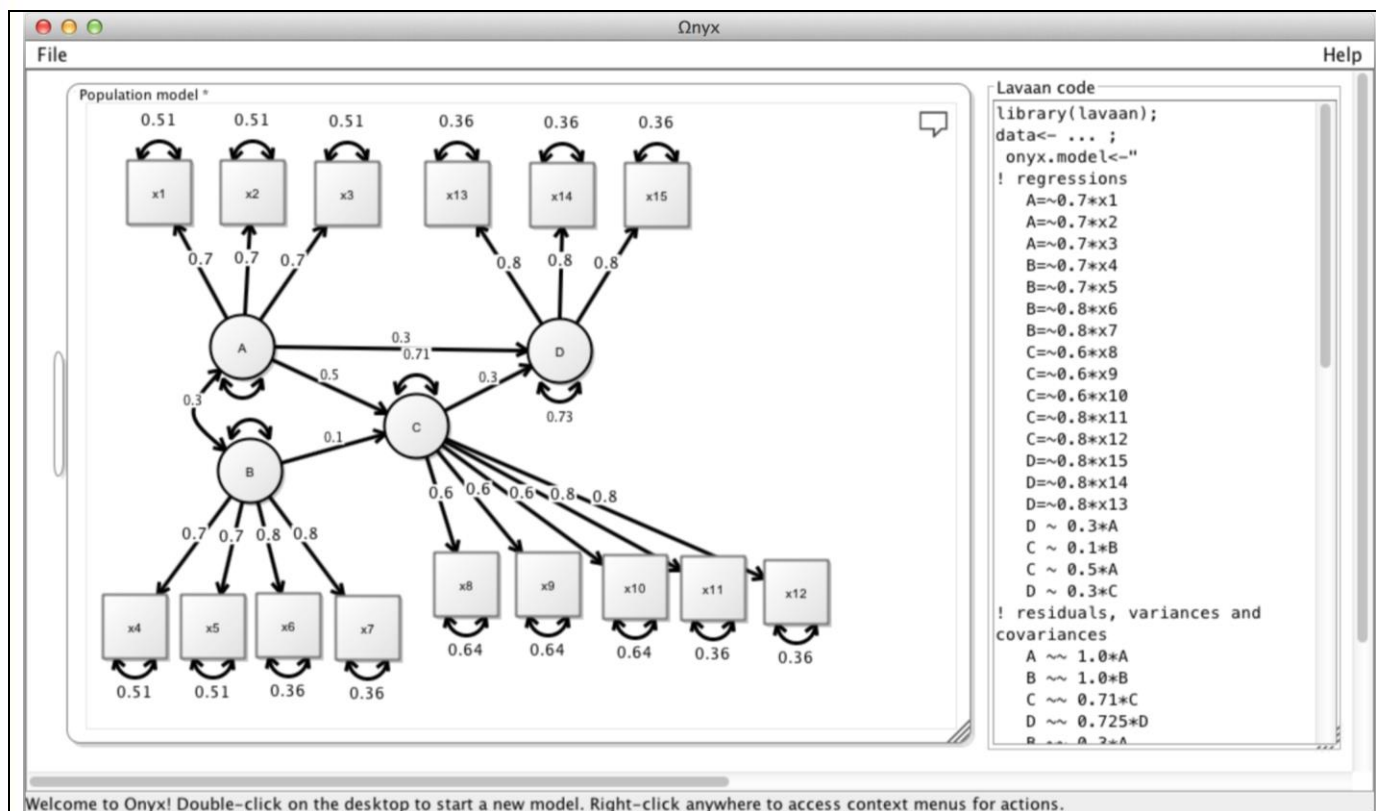

Figure 2. Defining the Population Model with Ωnyx

Researchers should then decide on the number of indicators that they will use to measure each construct. For simplicity, all four constructs in our running example are reflectively specified, though formative specifications are also possible. In our example, construct A is measured with three indicators, B measured with four indicators, C measured with five indicators, and D measured with three indicators. All these are essentially congeneric, which entails each item loading on a single construct, with potentially different loadings across items; measurement errors are assumed to be uncorrelated. Both of these conditions, however, can be relaxed if needed. All variables in the model, latent or manifest, have zero means and unit variance. Using a model where all variables are standardized is recommended because the PLS results are always standardized and comparing estimates to the population values is therefore simplified.

As a result, the loadings for each item need to be specified as well. In our example, we have set all item loadings for construct A at 0.70, half of those for construct B at 0.70 and half at 0.80, three loadings for items measuring construct C at 0.60 and the remaining two at 0.80, and all three loadings for construct D at 0.80. Because all items are here specified as having unit variance, the error variances are set at the square root of one minus the square of the respective loading. Note that specifying all latent and observed variables with unit variance essentially entails a model with standardized parameters, which is comparable to the way most research results are reported.

After the population model has been fully specified, the next step is to draw repeated samples from the population so that those samples reflect the population model just defined and any distributional considerations that researchers are interested in analyzing. The case of data drawn from a multivariate normal distribution is more straightforward and so we discuss it first. In addition, this scenario lays down some of the groundwork that will be required later. After this first example, we discuss one way in which non-normal data can be generated for analysis.

We now export the generated model from Ωnyx to R. The code can be either copy-pasted, one part at a time to the R window, or combined as one text file that is then run with R (the recommended approach because it makes it easier for researchers to save their work for later review and to modify the code). The model syntax is shown below and we assign it to an object named "model".

```
model<-"
! regressions
A =~ 0.7*x1
A =~ 0.7*x2
A =~ 0.7*x3
B =~ 0.7*x4
B =~ 0.7*x5
B =~ 0.8*x6
B =~ 0.8*x7
C =~ 0.6*x8
C =~ 0.6*x9
C =~ 0.6*x10
C =~ 0.8*x11
C =~ 0.8*x12
D =~ 0.8*x13
D =~ 0.8*x14
D =~ 0.8*x15
D ~ 0.3*A
C ~ 0.1*B
C ~ 0.5*A
D ~ 0.3*C
! residuals, variances and covariances
A ~~ 1.0*A
B ~~ 1.0*B
C ~~ 0.71*C
D ~~ 0.725*D
B ~~ 0.3*A
x1 ~~ 0.51*x1
x2 ~~ 0.51*x2
x3 ~~ 0.51*x3
x4 ~~ 0.51*x4
x5 ~~ 0.51*x5
x6 ~~ 0.36*x6
x7 ~~ 0.36*x7
```

```
x8  ~~ 0.64*x8
x9  ~~ 0.64*x9
x10 ~~ 0.64*x10
x11 ~~ 0.36*x11
x12 ~~ 0.36*x12
x13 ~~ 0.36*x13
x14 ~~ 0.36*x14
x15 ~~ 0.36*x15"
```

The model syntax is similar to the Mplus model syntax and is explained in an paper that also contains multiple examples (Rosseel, 2012). The =~ denotes factor loadings of observed variables on the latent variables, and ~ denotes regression relationships between latent variables. Both are technically regression coefficients, but the factor loadings are marked differently because they also define latent variables (a formative latent variable would be defined similarly but not including an observed variable on the right side of the =~ relation and then specifying regression paths from the indicators toward the latent variable). The ~~ symbol is used for residual variances and covariances of endogenous variables and variances and covariances of exogenous variables. In our example, we specify the variances of A and B and the covariance between A and B. Because C and D are endogenous variables, the ~~ symbol defines their residual variances (i.e., latent variable level error terms).

Step 2: Monte Carlo simulation

The next step is to use the model in a Monte Carlo simulation. We start by loading the matrixpls package and use the matrixpls.sim function to perform a Monte Carlo study. This function has three mandatory parameters. The first parameter is the number of Monte Carlo replications, for which we recommend a value of at least 500 (note that matrixpls makes use of multiple cores in computers equipped with them, and thus the number of replications can be increased markedly if so desired). The next parameter is the population model. If the tested model is not defined in a separate parameter, the population model will also be used as the estimated model; a separate tested model would be indicated when researchers are interested in analyzing, for example, the consequences of employing an analysis model that is misspecified when compared to the population one. The final mandatory parameter is the sample size. The result will be assigned into an object named "output".

The two additional parameters that are highly recommended are multicore (see above), which enables parallel processing and takes advantage of contemporary multicore processors, and completeRep, which discards and re-runs failed (i.e., non-converged) estimation results so that the final number of replications is always guaranteed to be exactly the number specified.

```
library(matrixpls)

output <- matrixpls.sim(1000, model, n=100, multicore = TRUE, completeRep = TRUE)
```

More-complex setups are also possible. For example, generating non-normal data can be done by specifying the skewness and kurtosis values for each indicator variable, as shown below[1]:

```
distributions <- mapply(bindDist,
                  skewness = c(0.50, 0.50, 0.50, 0.50, 0.50,
                               0.50, 0.50, 0.50, 0.75, 0.75,
                               0.75, 0.75, 0.75, 0.75, 0.75),
                  kurtosis = c(1, 1, 1, 1, 1,
                               1, 1, 1, 1.50, 1.50,
                               1.50, 1.50, 1.50, 1.50, 1.50))

output <- matrixpls.sim(1000, model, n=100, indDist = distributions,
                        multicore = TRUE, completeRep = TRUE)
```

Both these examples generate 1000 samples of data from the population model and estimate the model parameters with each sample using the PLS algorithm with the default settings, which are to use mode A outer weighting scheme (typically used for reflective indicators) for the outer estimation and path scheme, which is the default for

---

[1] The full capabilities of the simsem package, which is used for data generation and analysis by matrixpls, are documented on the simsem website at http://simsem.org, which includes multiple ready-to-run examples.

also many other PLS applications, for inner weighting. Standard errors are estimated with 500 bootstrap replications and 95 percent confidence intervals are estimated using the normal approximation. In total, this setup estimates 501,000 PLS models, which takes about 15 minutes on a contemporary multi-core processor computer.

Step 3: Interpretation of the Results

After the Monte Carlo replications have been completed, the next step is to interpret the results. These are printed out by applying the summary function on the output object

```
summary(output)
```

This produces the following output

```
========= Fit Indices Cutoffs ============
          Alpha
Fit Indices   0.1   0.05  0.01 0.001 Mean   SD
      srmr 0.089 0.092 0.123 0.154 0.08 0.01
========= Parameter Estimates and Standard Errors ============
Called from: summaryParam(object, digits = digits)
Browse(1)>
     Estimate.Average Estimate.SD Average.SE Power..Not.equal.0. Average.Param Average.Bias Coverage
C~A            0.398       0.089      0.085               0.973           0.5       -0.102    0.789
D~A            0.253       0.099      0.100               0.705           0.3       -0.047    0.933
C~B            0.139       0.091      0.110               0.300           0.1        0.039    0.927
D~C            0.283       0.098      0.099               0.779           0.3       -0.017    0.945
A=~x1          0.814       0.097      0.100               1.000           0.7        0.114    0.735
A=~x2          0.801       0.100      0.101               1.000           0.7        0.101    0.802
A=~x3          0.805       0.101      0.101               0.999           0.7        0.105    0.770
B=~x4          0.759       0.161      0.199               0.919           0.7        0.059    0.906
B=~x5          0.760       0.166      0.201               0.936           0.7        0.060    0.920
B=~x6          0.824       0.148      0.193               0.949           0.8        0.024    0.951
B=~x7          0.825       0.150      0.193               0.951           0.8        0.025    0.952
C=~x8          0.684       0.106      0.109               0.998           0.6        0.084    0.865
C=~x9          0.696       0.102      0.108               0.999           0.6        0.096    0.836
C=~x10         0.690       0.109      0.109               0.999           0.6        0.090    0.834
C=~x11         0.834       0.090      0.089               1.000           0.8        0.034    0.917
C=~x12         0.834       0.091      0.089               1.000           0.8        0.034    0.906
D=~x13         0.866       0.087      0.088               1.000           0.8        0.066    0.862
D=~x14         0.869       0.087      0.088               1.000           0.8        0.069    0.859
D=~x15         0.871       0.087      0.088               1.000           0.8        0.071    0.845
================== Replications ====================
Number of replications = 1000
Number of converged replications = 1000
Number of nonconverged replications:
   1. Nonconvergent Results = 0
   2. Nonconvergent results from multiple imputation = 0
   3. At least one SE were negative or NA = 0
   4. At least one variance estimates were negative = 0
   5. At least one correlation estimates were greater than 1 or less than -1 = 0
```

The first section of the output contains the distributions of various model fit indices that were calculated during the estimation. With default settings, simsem prints out only a small selection of most commonly used SEM fit indices of which only standardized root mean squared residual (SRMR) is available from *matrixpls* (Rönkkö, 2013). Because there is limited evidence of the usefulness of these indices in the PLS context (Evermann & Tate, 2010 Rönkkö & Evermann, 2013), and because they fall outside the scope of our topic of power analysis, we do not discuss them any further.

The second section contains information about the parameter estimates. The estimates are labeled with the same Lavaan syntax (Rosseel, 2012) that was used to define the population model. The first column "Estimate.Average" shows the mean of all estimates and the "Estimate.SD" shows the standard deviation of the estimates and indicates their variability around the mean. "Average.SE" is the average estimate of standard error. Because the standard error is the estimate of the standard deviation, these values should be close to the actual standard errors to be unbiased.

The fourth column "Power..Not.equal.0" is that which is most relevant for our purposes. This column contains the results of the power analysis. The significance tests that are used in the power analysis are based on comparing the ratio of the parameter estimate to its standard error with the *t*-distribution. Although this practice has been challenged lately (Rönkkö & Evermann, 2013), the field currently lacks better alternatives. The statistic in the column

tells us the relative frequency that the p-value from this test is below .05, which provides an estimate of statistical power.

The final three columns provide information that can also be useful. "Average.Param" is the population value of the parameter and "Average.Bias" is the bias of the estimates, which is the difference between the mean estimate and the population value (the term "average" is used here because simsem supports also multigroup analyses). "Coverage" is the coverage value of the confidence intervals, which tells how frequently the population value of the parameter falls within the estimated confidence limits. Confidence intervals are sometimes suggested as a potential alternative to null-hypothesis significance testing (e.g., Henseler, Ringle, & Sinkovics, 2009), but our results—similarly to earlier results by Lu, Kwan, Thomas, and Cedzynski (2011)—show that, at least for the confidence intervals calculated using a normal approximation, the coverage is often not very close to the nominal 95 percent value. Nevertheless, a recent study by Henseler and colleagues (2014; see also McIntosh, Edwards, & Antonakis, 2014) provides initial evidence of the usefulness of more-complex bias-corrected and accelerated confidence intervals. Although these confidence intervals were not used in our tutorial, they can be obtained by adding the option `citype = "bca"` to the simulation command.

From the perspective of research design and planning, these results are valuable for two reasons. First, they provide researchers with information about what to expect when conducting their study. In particular, these results indicate the average estimates for each main parameter of interest and their standard errors, which provide a measure of the expected variability of the estimates and the range of findings that are likely from a population model such as the one specified by the researcher. For example, for a population model such as the one specified in Figure 2 and a sample size of one hundred as used here, the expected average path value for the A -> C path would be 0.398, whereas approximately sixty eight percent of the estimates would fall between 0.309 and 0.487 (that is, within one standard deviation from the mean).

Second, and specifically with regards to statistical power, the results provide an estimate of power to detect the specified effects at the chosen sample size. In this case, a sample size of one hundred, as used in the simulation, would have a power of 0.973 to detect the A -> C path, but a power of only 0.300 to detect the B -> C path, well below commonly accepted standards. As we note before, findings from underporwered studies, when significant, should be cautiously considered. These models can then be rerun with different combinations of effect and sample sizes in order to ascertain the likelihood of detecting underlying effects, which aids in both planning the research and interpreting the obtained results.

Comparison between Approaches

The simulation approach presented here can be easily adapted for use with CBSEM as well because it builds directly on simulation tools that were developed for CBSEM. This can be done by replacing the function from matrixpls.sim to just sim to run the same model with Lavaan (Rosseel, 2012). Additionally, two models need to be specified, the first model being the estimated model and the second the population model where all parameters are freed. This is similar to how Monte Carlo simulations are specified when using Mplus. Appendix A provides an example of how this kind of analysis is specified.

After running the same analysis by estimating the same model with CBSEM, we can compare the results between estimating power directly by using PLS in a Monte Carlo simulation, approximating this simulation with CBSEM simulation, and using power tables from the work of Cohen. Table 2 presents the power estimates for this analysis for PLS and CBSEM showing that the power estimates using PLS are substantially higher than the power estimates for CBSEM, and that neither of two approaches reaches the targeted power level of 0.80 but for one path, though PLS comes close for two other paths. However, this comparison is incomplete unless we also compare these results with power tables. Because these tables do not consider individual coefficients, but rather the overall explained variance, they must be applied on a per construct, rather than on a per path, basis. The population R^2 values for C and D are 0.29 and 0.27, translating to f^2 values of 0.41 and 0.37. These values suggest that it would be most appropriate to compare these results for the power tables of large effect (0.35) for a regression with two explanatory variables. In this scenario, the tables state that 30 observations would be sufficient to reach the target power level of 0.80.

Table 2: Comparison of Power Analysis Approaches		
	Analysis	
Parameter	PLS simulation	CBSEM simulation
A -> C (0.5)	0.973	0.965
B -> C (0.1)	0.300	0.135
A -> D (0.3)	0.705	0.543
C -> D (0.3)	0.779	0.504

Summary and Application

This tutorial proposes a straightforward approach to determining sample size and calculating statistical power for research employing PLS as the analytical technique of choice—one which does not rest on rules-of-thumb or on approaches developed for different statistical techniques. Furthermore, the steps necessary to perform these calculations are largely automated and do not require extensive additional work beyond what a researcher would need to perform to carry out the research in the first place.

The first step is for researchers to specify their research model of interest in the appropriate syntax. This can be accomplished either manually, following the conventions by Rosseel (2012), or by using a graphical tool, such as Ωnyx, which we employed here. Both alternatives would result in an identical outcome, and it is thus largely a matter of personal preference. The most important part of this step is specifying reference values for each parameter in the model. While the structure of the research model itself (which constructs are included, the relationships between those constructs, the number of indicators used to measure each construct, etc.) should have already been developed as part of the research project itself—and thus presents no additional effort beyond the need to translate the research model into the appropriate syntax—researchers also need to specify what they consider to be likely values for each parameter in the research model. These values are central to the determination of sample size because the expected power resulting from the Monte Carlo simulations is dependent on the particular values specified in the research model. It is always possible to create several different models with the same structure but with different values for certain parameters in order to ascertain the sensitivity of the results to particular values.

The second step involves choosing a desired distribution for the simulation and its subsequent execution. In the two examples above, we show how to employ a multivariate normal distribution—which is the default option in the package and thus requires no additional configuration—or the specification of a non-normal distribution for the observed variables by entering values for their skewness and kurtosis. It is also possible to employ a mixed approach by noting that a normally distributed variable has both kurtosis and skewness equal to zero. In this step, it is also possible to specify both the desired number of replications, where a higher number leads to more stable results, and the expected sample size. As before, the analyses can be run with a variety of sample sizes in a likely range in order to better ascertain the sensitivity of the results to this parameter. The statistical power reported for each parameter in the model can only be interpreted by reference to the particular value assigned to that parameter in the simulation and the chosen sample size.

V. RECOMMENDATIONS AND CONCLUSION

The procedure we outline above can be readily modified to conduct power analyses for any PLS model, for any number of constructs, indicators, sample size, inner and outer weighting schemes, or distributional assumptions. These procedures are computing intensive, and, thus, can take some time to complete. In the example we discuss here, obtaining one thousand converged replications, with five hundred bootstrap resamples each entails generating and analyzing more than half a million runs. That said, we believe the benefits of performing these calculations, both in terms of research planning and sample size determination, and in aiding to the interpretation and support of research findings, clearly outweigh a few minutes of processing time.

We therefore strongly recommend that researchers both conduct and report the results of power analyses as we discuss here. The following elements should be reported for each separate power analysis, which can be included in an appendix so as not to clutter the reporting of the main research:

1. The population model used to generate the data should be structurally identical to the model used later in the analysis (i.e., same constructs and relationships among them, same number of items, same assumptions as to loadings of indicators on each construct, and same cross-loadings and correlated measurement errors if any). The values used for each parameter in the simulation should be reported.
2. The individual sample size employed (in the example reviewed here, one hundred).
3. The number of converged replications on which the results are based, and the number of bootstrap replications obtained for each individual dataset.
4. Distributions used to generate the sample data.

5. Statistical power achieved for every parameter of interest.

Including the elements above would also facilitate the replication of the analysis by other researchers if so desired. The following sample write-up conveys all this information for the example used here (for the normal case), which also shows that the reporting of power analyses need not be onerous:

> *Statistical power analyses were conducted in order to determine the sample size necessary in order to achieve a desired level of power to detect an existing effect. The population model used for the power analysis was structurally identical to the one shown in Figure 2. Standardized loadings for the items measuring A were 0.7, for half of those measuring B, 0.70, and for the other half, 0.8. The first three items measuring C loaded at 0.6, whereas the remaining two had standardized loadings set at 0.8. All items measuring D loaded at 0.8. We drew all variables, latent and manifest, from a multivariate normal distribution with zero mean and unit variance. All items loaded solely on their intended construct, with no correlated errors assumed. The path from A to C was set at 0.50, that from B to C at 0.10, that from A to D at 0.30, and that from C to D at 0.30. The individual sample size was one hundred, and the results are based on one thousand converged replications, with five hundred bootstrap resamples each. These results indicate a power of 0.973 for the path between A and C, 0.300 for the path between B and C, 0.704 for the path between A and D, and 0.779 for the path between C and D.*

The same procedure we outline here for the assessment of statistical power could be used, for example, to test false positive rates. In this scenario, researchers would be interested in assessing the extent to which a PLS analysis of a particular model and distribution could be expected to result in estimates that are significantly different from zero when they are indeed zero in the population. Such analysis might alleviate the concerns of possible false positives due to violation of the assumptions of the t-test (Rönkkö & Evermann, 2013). This is but one of many different analyses that researchers can conduct by being able to easily specify and execute simulations using the packages and code provided here. Additional scenarios could include the presence of cross-loadings, correlated measurement errors, and others.

In this paper, we argue for the importance of conducting a priori power analyses when employing PLS as the statistical technique of choice. The value of doing so rests not only on the ability to provide strongly grounded evidence for the appropriateness of the research design underlying an empirical study, but also in the information it provides researchers to aid in the planning and data collection stages. In order to facilitate the conduct of these analyses in the future, we provide here a tutorial on the usage of the matrixpls package for this purpose. In conjunction with the Ωnyx application—which greatly simplifies the specification of the population models, though those could be also done by hand using the Lavaan syntax—this allows researchers to quickly and easily specify and test a variety of population models that reflect the range of conditions or scenarios of interest in a research area. We also provide guidance in interpreting and reporting the results from these analyses. Through all these, we contribute to the improvement of research practice in our field by facilitating the more nuanced use of one of the most popular statistical techniques.

ACKNOWLEDGMENTS

REFERENCES

Editor's Note: The following reference list contains hyperlinks to World Wide Web pages. Readers who have the ability to access the Web directly from their word processor or are reading the paper on the Web, can gain direct access to these linked references. Readers are warned, however, that:
1. These links existed as of the date of publication but are not guaranteed to be working thereafter.
2. The contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. The author(s) of the Web pages, not AIS, is (are) responsible for the accuracy of their content.
The author(s) of this article, not AIS, is (are) responsible for the accuracy of the URL and version information.

Aguirre-Urreta, M., Marakas, G., & Ellis, M. (2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *The DATA BASE for Advances in Information Systems*, *44*(4), 11-43.

Ang, S., & Straub, D. (1998). Production and transaction economies and IS outsourcing: A study of the U.S. banking industry. *MIS Quarterly*, *22*(4), 535–552.

Baroudi, J., & Orlikowski, W. (1989). The problem of statistical power in MIS research. *MIS Quarterly*, *13*(1), 87–106.

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

Chin, W., & Newsted, P. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–341). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.

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

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* London: Lawrence Erlbaum Associates.

Enns, H., Huff, S., & Higgins, C. (2003). CIO lateral influence behaviors: Gaining peers' commitment to strategic information systems. *MIS Quarterly*, *27*(1), 155–176.

Evermann, J., & Tate, M. (2010). Testing models or fitting models? Identifying model misspecification in PLS. In *ICIS 2010 Proceedings.* Retrieved from http://aisel.aisnet.org/icis2010_submissions/21

Furneaux, B., & Wade, M. (2011). An exploration of organizational level information systems discontinuance intentions. *MIS Quarterly*, *35*(3), 573–598.

Gefen, D., Rigdon, E., & Straub, D. (2011). An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly*, *35*(2), iii–xiv.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J. J., Hair, J. F., Hult, G. T. M., & Calantone, R. J. (2014). Common beliefs and reality about PLS comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*(2), 182–209.

Henseler, J., Ringle, C., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, *20*(1), 277–319.

Hollenbeck, J., DeRue, D., & Mannor, M. (2006). Statistical power and parameter stability when subjects are few and tests are many: Comment on Peterson, Smith, Martorana, and Owens (2003). *Journal of Applied Psychology*, *91*(1), 1–5.

Ioannidis, J. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648.

Jiang, Z., & Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*, *31*(3), 475–500.

Kaplan, D. (1995). Statistical power in structural equation modeling. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 100–117). Thousand Oaks, CA: Sage Publications.

Klein, R., & Rai, A. (2009). Interfirm strategic information flows in logistic supply chain relationships. *MIS Quarterly*, *33*(4), 735–762.

Lewis, W., Agarwal, R., & Sambamurthy, V. (2003). Sources of influence on beliefs about information technology use: An empirical study on knowledge workers. *MIS Quarterly*, *27*(4), 657–678.

Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, *31*(1), 59–87.

Lu, I., Kwan, E., Thomas, D., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, *28*(3), 258–268.

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

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

Marcoulides, G., & Saunders, C. (2006). PLS: A silver bullet? *MIS Quarterly*, *30*(2), iii–ix.

Maxwell, S. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*(2), 147–163.

McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods*, *17*(2), 210–251.

Muthén, L., & Muthén, B. (1998). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

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

von Oertzen, T., Brandmaier, A. M., & Tsang, S. (2015). Structural equation modeling with Ωnyx. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 148–161.

Peng, D., & 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.

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2013). *Simsem: SIMulated structural equation modeling*. Retrieved from http://CRAN.R-project.org/package=simsem

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of research in marketing*, *26*(4), 332–344.

Ringle, C., Sarstedt, M., & Straub, D. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, *36*(1), iii–xiv.

Rönkkö, M. (2014). The effects of chance correlations on partial least squares path modeling. *Organizational Research Methods*, *17*(2), 164–181.

Rönkkö, M. (2014). *Matrixpls: Matrix-based partial least squares estimation*. Retrieved from http://cran.r-project.org/web/packages/matrixpls/index.html

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425–448.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Siponen, P., & Vance, A. (2010). Neutralization: New insights into the problem of employee information systems security policy violations. *MIS Quarterly*, *34*(3), 487–502.

Venkatesh, V., & Morris, M. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behaviors. *MIS Quarterly*, *24*(1), 115–139.

Wells, J., Valacich, J., & Hess, T. (2011). What signals are you sending? How website quality influences perceptions of product quality and purchase intentions. *MIS Quarterly*, *35*(2), 373–396.

Wolf, E., Harrington, K., Clark, S., & Miller, M. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*, 913–934.

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspectives on Psychological Science*, *4*(3), 294–298.

Zumbo, B., & Hubley, A. (1998). Misconceptions concerning prospective and retrospective power. *Journal of the Royal Statistical Society, Series D*, *47*(2), 385–388.

APPENDIX A: FULL CODE FOR THE EXAMPLE

```
library(matrixpls)

model<-"! regressions
A=~0.7*x1
A=~0.7*x2
A=~0.7*x3
B=~0.7*x4
B=~0.7*x5
B=~0.8*x6
B=~0.8*x7
C=~0.6*x8
C=~0.6*x9
C=~0.6*x10
C=~0.8*x11
C=~0.8*x12
D=~0.8*x13
D=~0.8*x14
D=~0.8*x15
```

```
D ~ 0.3*A
C ~ 0.1*B
C ~ 0.5*A
D ~ 0.3*C
! residuals, variances and covariances
A ~~ 1.0*A
B ~~ 1.0*B
C ~~ 0.71*C
D ~~ 0.725*D
B ~~ 0.3*A
x1 ~~ 0.51*x1
x2 ~~ 0.51*x2
x3 ~~ 0.51*x3
x4 ~~ 0.51*x4
x5 ~~ 0.51*x5
x6 ~~ 0.36*x6
x7 ~~ 0.36*x7
x8 ~~ 0.64*x8
x9 ~~ 0.64*x9
x10 ~~ 0.64*x10
x11 ~~ 0.36*x11
x12 ~~ 0.36*x12
x13 ~~ 0.36*x13
x14 ~~ 0.36*x14
x15 ~~ 0.36*x15"

# Normal data

output <- matrixpls.sim(1000, model, n=100, # Basic parameters
                        multicore = TRUE,   # Additional parameters
                        completeRep = TRUE)

summary(output)

# Non-normal data

distributions <- mapply(bindDist,
                        skewness = c(0.50, 0.50, 0.50, 0.50, 0.50,
                                     0.50, 0.50, 0.50, 0.75, 0.75,
                                     0.75, 0.75, 0.75, 0.75, 0.75),
                        kurtosis = c(1, 1, 1, 1, 1,
                                     1, 1, 1, 1.50, 1.50,
                                     1.50, 1.50, 1.50, 1.50, 1.50))

nonNorm <- matrixpls.sim(1000, model, n=100,         # Basic parameters
                         indDist = distributions, # Indicator distr
                         multicore = TRUE,         # Additional params
                         completeRep = TRUE)

summary(nonNorm)

# CBSEM, normal data

cbsem <- sim(nRep = 1000,
             model = gsub("(0-9.)+\\*","",model), # Free all params
             generate = model, n=100, # Basic parameters
             lavaanfun = "sem", # Do a SEM analysis
             multicore = TRUE,  # Additional parameters
             completeRep = TRUE)
```

```
summary(cbsem)
```

APPENDIX B: FULL CODE FOR BOOTSTRAPPING CORRELATION

```
library(parallel)

observations <- 17
boot.observations <- 500
boot.reps <- 1000

options(mc.cores = detectCores())

tstats <- mclapply(1:1000, function(rep){
  data <- cbind(rnorm(17), rnorm(17))
  boot.reps <- sapply(1:1000, function(x){
    boot.sample <- data(sample(1:observations, boot.observations, replace = TRUE),)
    cor(boot.sample(,1), boot.sample(,2))
    }, simplify = TRUE)
  cor(data(,1), data(,2))/sd(boot.reps)
  })

tstats <- unlist(tstats)
sum(abs(tstats)>1.96)
```

ABOUT THE AUTHORS

Miguel I. Aguirre-Urreta is currently an Assistant Professor in the Area of Information Systems and Quantitative Sciences, Jerry S. Rawls College of Business Administration, at Texas Tech University. He received his PhD in Information Systems from the University of Kansas and his MBA in Information Systems and Finance from the Kelley School of Business at Indiana University. His undergraduate studies, in Public Accounting, were completed at the Universidad de Buenos Aires in Argentina. His work has appeared or is forthcoming at *MIS Quarterly*, *Information Systems Research*, *Communications of the Association for Information Systems*, and *The DATA BASE for Advances in Information Systems*, among others. He is a Senior Editor at *The DATA BASE for Advances in Information Systems* and the incoming Chair Elect for the AIS Special Interest Group in Human-Computer Interaction (SIGHCI). His research interests are in the areas of quantitative research methods, computer self-efficacy, technology acceptance and diffusion, human-computer interaction, and business analytics.

Mikko Rönkkö is a post-doctoral researcher at Aalto University, School of Science, Department of Industrial Engineering and Management, Institute of Strategy and Venturing, where he also received his PhD. His research interests are in the areas of quantitative research methods and strategy and entrepreneurship. He is the author of the matrixpls R package.