

JYX



JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Aguirre-Urreta, Miguel I.; Rönkkö, Mikko; Hu, Jiang

Title: Polynomial Regression and Measurement Error : Implications for Information Systems Research

Year: 2020

Version: Accepted version (Final draft)

Copyright: © Association for Computing Machinery (ACM), 2020

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Aguirre-Urreta, M. I., Rönkkö, M., & Hu, J. (2020). Polynomial Regression and Measurement Error : Implications for Information Systems Research. Data Base for Advances in Information Systems, 51(3), 55-80. <https://doi.org/10.1145/3410977.3410981>

Polynomial Regression and Measurement Error: Implications for Information Systems Research

Miguel I. Aguirre-Urreta
Florida International University

Mikko Rönkkö
University of Jyväskylä

Jiang Hu
University of Texas – Arlington

Acknowledgments

The work by Mikko Rönkkö was supported by a grant from the Academy of Finland (grant 311309).

Abstract

Many of the phenomena of interest in information systems (IS) research are nonlinear, and it has consequently been recognized that by applying linear statistical models (e.g., linear regression), we may ignore important aspects of these phenomena. To address this issue, IS researchers are increasingly applying nonlinear models to their datasets. One popular analytical technique for the modeling and analysis of nonlinear relationships is polynomial regression, which in its simplest form fits a “U-shaped” curve to the data. However, the use of polynomial regression can be problematic when the independent variables are contaminated with measurement error and the implications of error can be more severe than in linear models. In this research, we discuss a number of techniques that can be used for modeling polynomial relationships while simultaneously taking measurement error into account and examine their performance by using a simulation study. In addition, we also discuss the use of marginal and response surface plots as interpretational aides when evaluating the results of polynomial models and showcase their use through a practical example using a well-known dataset. Our results clearly indicate that the use of a linear regression analysis for this kind of models is problematic, and we provide a set of recommendations for future IS research practice.

Keywords: Polynomial regression, Measurement error, Attenuation, Nonlinear SEM, Latent variables.

Introduction

Many of the phenomena of interest in systems (IS) research are nonlinear, and it has consequently been recognized that by applying linear statistical models (e.g., linear regression), we may ignore important aspects of these phenomena. A few recent exemplars published in the top journals of the discipline include the work of Lankton, McKnight, Wright and Thatcher (2016), who combined expectation disconfirmation theory and trust theory together with polynomial modeling in order to propose and validate a nonlinear model of trusting intention as a function of preusage technology trust expectations and postusage modified technology trust beliefs and nonlinear functions of those (e.g., multiplicative and powered terms); the work of Moody, Lowry and Galletta (2017), who modeled trusting intentions as a function of competence (trusting) beliefs, incompetence (distrusting) beliefs, and their multiplicative and powered terms; or the work of Venkatesh and Goyal (2010), who examined the effects of positive and negative expectation disconfirmation on behavioral intention to continue using a system, with a nonlinear specification that also included multiplicative and powered terms of the predictors.

The most popular analytical technique in this kind of studies is polynomial regression, which includes multiplicative and higher-order (e.g., powered) terms of the first-order predictors as part of the model (see, for example, the work of Lankton et al., 2016). In the simplest possible scenario such model fits a “U-shaped” curve to the data, but also more complex functions are possible and have been used in past IS research¹. To assess how polynomial regression has been used in IS research we did a keyword search using the terms ‘polynomial’ or ‘quadratic’ of the premier journals in the discipline (i.e., MIS Quarterly, Information Systems Research, Journal of Management Information Systems, and the Journal of the Association for Information Systems) in the period 2005-2016. Retaining only the studies that applied polynomial regression produced a list of 34 articles, which are summarized in Table 1².

Insert Table 1 About Here

As Table 1 clearly shows, regression on observed variables — either sumscores of individual indicators or single-item measures — is by far the most common technique for estimating polynomial models in the IS discipline. Given that observed indicators nearly always contain some measurement error, a sumscore of the indicators would also not be a perfectly reliable proxy for the construct of interest. While both the methodological literature (e.g., Edwards & Parry, 1993) and applied IS research (e.g., Venkatesh & Goyal, 2010) have noted that polynomial regression assumes that the independent variables are measured without error, and that violating this assumption will lead to biased estimates, the consequences of violating this assumption have not been carefully examined in IS research. This is somewhat surprising given that the biasing effects of measurement error on regression estimates are well-known (Goodhue, Lewis, & Thompson, 2017) and a number of procedures to address the issue in a linear model context exist (Dijkstra & Henseler, 2015; Rönkkö, McIntosh, & Aguirre-Urreta, 2016). What appears less well understood is that this effect is much greater in the case of polynomial regression models and that measurement error can influence not only the magnitude (and possibly sign) but also the shape of the relationship. This problem

is further compounded by the fact that the interpretation of polynomial regression results in the IS literature is frequently limited to an examination of the sign and significance of the model coefficients. However, unlike the linear case, complex nonlinear effects cannot be interpreted this way because the effect is not constant over the range of the independent variables.

Therefore, we have two main goals for this research. First, we analyze the performance of a variety of analytical techniques that do take measurement error into account and compare their parameter estimation accuracy. Specifically, we focus on the biasing effects of measurement error on estimates of the relationship between the multiplicative and powered terms in a polynomial regression equation and the dependent variable of interest, and how various techniques that can be used to estimate these models perform relative to one another on key outcomes of interest (estimate accuracy, Type I error, statistical power, etc.) and across various simulation factors (sample size, number of indicators, loading strength, etc.). Second, we discuss how marginal and response surface plots can and should be used to aid in the interpretation of polynomial regression results, demonstrating their use through a practical application with a well-known dataset.

Measurement Error and Polynomial Regression

The effects of measurement error in linear regression models differ depending on whether the dependent or independent variables are contaminated with error. In the case of the dependent variable, the effect of adding random noise is straightforward: measurement error does not bias the resulting regression estimates, but because error increases the variance of the dependent variable, it does attenuate (that is, biases toward zero) the multiple correlation between the predictors and criterion (e.g., the R in R^2) and influences standardized estimates the same way. Measurement error also decreases the precision of the regression estimates, consequently increasing their standard errors (Ree & Carretta, 2006), and thus decreases statistical power.

There are two different scenarios that need to be considered when ascertaining the effects of measurement error in the independent variables in a research model. First, in the case of one independent variable, or an independent variable that is uncorrelated with other independent variables, measurement error produces a bias toward zero due to attenuation (e.g., Wooldridge, 2013, p. 322). Second, in the case of multiple independent variables that are correlated, measurement error in one independent variable leads to bias in all regression coefficients but determining the magnitude or even the direction of bias is difficult (e.g., Wooldridge, 2013, p. 322). This is highlighted in a recent study by Goodhue et al. (2017), who demonstrated a scenario where measurement error leads to false positive findings.

When multiplicative and powered terms are included, the effects of measurement error are exacerbated because the reliability of the multiplicative and powered terms is a fraction of the reliability of the first-order predictors. Even if the reliabilities of the variables that form the product exceed the 0.7 cutoff that is often used as a minimum threshold that must be reached in IS research³, the reliability of multiplicative and powered terms may not reach the same standard (Dalal & Zickar, 2012). Consequently, the regression estimates capturing the effects of multiplicative and powered terms may be substantially more biased than what would be expected based just on the reliabilities of the original variables. Clearly, even if the reliabilities of the first-order components were adequate, this does not necessarily provide sufficient protection from the negative consequences of measurement error for the entire model.

To see why this is the case, consider the formula for the reliability of a product of two variables (Dalal & Zickar, 2012):

$$\rho_{XZ} = \frac{\rho_X \rho_Z + r_{XZ}^2}{1 + r_{XZ}^2} \quad [1]$$

where ρ_x and ρ_z are the reliabilities of the first-order components, and r_{XZ}^2 is the square of the correlation between the first order components. Based on Equation 1, it is clear that the reliability is at minimum the product of the two reliabilities when the variables are uncorrelated and improves as the correlation increases. To give an example, the reliability of a multiplicative term of two first-order components that are correlated at 0.3 (a medium-sized correlation) and each of which exhibits a reliability of 0.7 (which would be considered adequate) is only 0.532. Commonly applied validation guidelines (e.g., Hair, Black, Babin, & Anderson, 2010, Chapter 3; MacKenzie, Podsakoff, & Podsakoff, 2011; Straub, Boudreau, & Gefen, 2004) would discourage researchers from interpreting results when the involved explanatory variable exhibits such low levels of reliability, yet the reliability of the higher-order terms

was ignored in all the IS articles that we reviewed. Even if the reliabilities of both first-order components had been 0.80, the reliability of the multiplicative term (0.67 in this case) would still be below the commonly-used 0.7 cutoff.

In turn, the reliability of a quadratic term is the square of the reliability of the original variable (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007):

$$\rho_{X^2} = (\rho_X)^2 \quad [2]$$

In this case, it is easy to see that the reliability of the quadratic product of the original variable can only be a fraction of the original reliability; when the original reliability is, for example 0.7 or 0.8, both of which would be considered adequate, the reliability of the quadratic term would only be 0.49 or 0.64, respectively, none of which would be considered adequate. Moreover, these low reliabilities would lead to severe downward bias in the estimates of their relationships to the dependent variable, as previously noted.

To summarize, though measurement error of the independent variables is an issue for any application of regression, its consequences are exacerbated in the presence of multiplicative and powered terms. This problem occurs because the product and quadratic terms of unreliable variables are even less reliable, which in turn leads to biased estimates and loss of statistical power to detect the effects of interest. In what follows, we outline a few alternative techniques that can be used to address the problem of measurement error, and then conduct a simulation study to understand their comparative performance under a variety of conditions.

Measurement Error and Polynomial Regression

In order to better understand the consequences of measurement error for polynomial research designs and the relative effectiveness of alternatives to ordinary least squares (OLS) regression, which is the most commonly used technique for estimating polynomial models, we conducted an extensive set of Monte Carlo simulations, as follows. The structural portion of the population model included two first-order independent variables, their interaction, and quadratic terms, as commonly done in the IS studies reviewed above:

$$Y = \gamma_1 X + \gamma_2 Z + \gamma_3 XZ + \gamma_4 X^2 + \gamma_5 Z^2 + \varepsilon \quad [3]$$

The correlation between the first-order independent variables (X, Z) was 0.1, 0.3, or 0.5 (which correspond to small, medium, or large correlation; Cohen, 1992). The values for the γ_1 and γ_2 parameters were initially set at to 0.5 and 0.6 and then scaled so that the base variance explained by X and Z alone without the higher-order terms would be 30%. The parameters γ_3 , γ_4 , and γ_5 always received the same value, which was determined to achieve the desired effect size f^2 for these higher-order terms, varied as an experimental condition taking values 0, 0.02, 0.15, or 0.35 (which correspond to no effect, and small, medium, or large effects; Cohen, 1992). Each latent variable was measured with 3, 5, or 7 indicators that loaded on their respective latent variables at 0.7, 0.8, or 0.9. The sample size was 100, 300, or 500. This resulted in 324 different simulation conditions: first-order correlation (3 levels) x regression effect size (4 levels) x number of indicators (3 levels) x loadings (3 levels) x sample size (3 levels). All variables were standard multivariate normal (mean = 0, SD = 1). Within each of these conditions we replicated the analysis 1,000 times. We then analyzed the data with five different approaches (OLS, PLSc, DR, EXT, and LMS), discussed next. All data generation was performed using custom code written for the *R Statistical Environment* (R Core Team, 2016, version 3.3.3).

Ordinary Least Squares (OLS). This is the current approach employed in the IS literature for the analysis of polynomial regression models and is included here to demonstrate the negative consequences of not considering the effects of measurement error on the estimated regression coefficients, as well as to provide a benchmark against which the other considered techniques can be evaluated. This approach started by summing the indicators measuring X, Z, and Y as scale scores that were then mean-centered and standardized, following the practice in the reviewed IS research⁴. Third, the scores for X and Z were used to create scores for the interaction and quadratic terms. Finally, all these were entered into a linear regression model predicting the composite score representing the Y construct. Significance of the estimates was assessed by means of the conventional t-tests as provided by default by the *lm* R function used for this analysis.

Consistent Partial Least Squares (PLSc). This approach was recently introduced to the IS discipline by the work of Dijkstra and Henseler (2015). The traditional PLS analysis is simply a regression with scale scores constructed as weighted sums of the indicators, and thus suffers from the measurement error problems explained earlier. The PLSc technique enhances the traditional PLS approach by applying a well-known correction for attenuation to the

scale score correlation matrix before regression estimation. In more detail, the PLSc algorithm works as follows (Dijkstra & Henseler, 2015, fig. 2). First, estimated scores for each of the latent variables are obtained from a traditional PLS analysis, which also provides a correlation matrix of the composites (which at this stage is inconsistent due to attenuation). Second, reliability estimates for the scores are calculated based on the PLS weights instead of the commonly used Composite Reliability index, which is consistent only if each indicator is weighted equally (i.e. unweighted sum) and individual indicator reliabilities are estimated consistently (Aguirre-Urreta, Marakas, & Ellis, 2013). Third, the correlation matrix of the composites is corrected using the classical correction for attenuation. Finally, the path coefficients between the constructs of interest are calculated by applying OLS regression to the corrected correlation matrix.

When applied to the particular case of polynomial regression, an additional set of adjustments is needed (see Dijkstra & Schermelleh-Engel, 2014). These involve using the variances and covariances of the independent variables to construct a disattenuated covariance matrix of the first-order, interaction, quadratic, and dependent variables, following the formulas expressed by Dijkstra and Schermelleh-Engel (2014; see also Brandt, Umbach, & Kelava, 2015, Appendix A). These adjustments require a normality assumption, and when this assumption holds, produce a consistent estimate of the covariance matrix of the independent variables (both first- and higher-order terms). After estimation, we used a bootstrap percentile confidence intervals (Aguirre-Urreta & Rönkkö, 2018; Dijkstra & Henseler, 2015) for statistical inference. To assess the statistical power of this approach, we converted the confidence intervals to hypothesis tests by checking whether zero was included in the 95% confidence interval⁵. This approach was implemented using the *matrixpls* R package (Rönkkö, 2017, version 1.0.4) and custom R code that added the disattenuated correlations of the quadratic and interaction terms to the scale score correlation matrix before regression estimation.

Disattenuated Regression (DR). The third approach is a variant of OLS regression that seeks to correct bias in the estimates that is due to the presence of measurement errors in the composites used to represent the constructs of interest in a path model, as discussed above. This approach is also known as errors-in-variables (EIV) regression (Culpepper, 2012; Fuller, 1987), and it has received some attention in recent methodological research (Devlieger & Rosseel, 2017) with encouraging results. This approach is otherwise identical to PLSc, but differs in that instead of using the PLS algorithm to set the indicator weights, the scale scores are calculated by simply summing the indicators and reliabilities are estimated with the Composite Reliability index, which is calculated from factor analysis results and is thus consistent (Aguirre-Urreta et al., 2013). The DR approach has a few advantages over PLSc, in that the approach is simpler, does not require special software, and avoids the computational issues related to PLS weights (Rönkkö et al., 2016). This approach was implemented in a manner similar to the PLSc approach just discussed, using *matrixpls* (Rönkkö, 2017, version 1.0.4).

Latent Moderated Structural Equations (LMS). This is the first technique, based purely on latent variables, that does not calculate scales scores as a part of the estimation procedure. The LMS approach, developed by Klein and Moosbrugger (2000), is a general latent interaction approach that implements a maximum likelihood (ML) estimation method, and is currently implemented in *MPlus* (Muthén & Muthén, 2004) and the R package *nlsem* (Umbach et al., 2015). Technical details about the development and implementation of the algorithm can be found in Klein and Moosbrugger (2000). The key insight that allowed the development of this approach is that the distribution of the dependent variable in a polynomial model is necessarily nonnormal, and if we assume that the measurement errors are normally distributed, information about the nonnormality of the indicators can be used to estimate the polynomial model. In practice, LMS represents the distribution of observed variables as a finite mixture of normal distributions, that is then fitted to the data using expectation maximization (EM) algorithm. Standard errors for the parameters of interest obtained by means of an information matrix, as is commonly done in ML estimation, and z tests are used for inference. This approach was implemented using *Mplus* (version 5.0). Because *Mplus* does not implement standardized estimates for LMS, we standardized the estimates using the technique presented by Brandt et al (2015) to make the results comparable across techniques.

Extended Unconstrained Approach (EXT). Another common technique for estimating a polynomial model in a latent variable context is to leverage the fact that products of indicator variables can be used as measures of the latent polynomial term. This approach to the modeling of nonlinear relationships with latent variables was originally introduced by Jaccard and Wan (1995), Ping (1996), and Jöreskog and Yang (1996), and later simplified and extended by Marsh et al (2004). This latter variation has the benefit of relaxing the complex constraints that were required to properly specify the relationships implied by the nonlinear terms. When both interaction and quadratic terms are simultaneously included, the extension by Moosbrugger et al (2009) needs to be considered. In this case, used in our study, error covariances of the nonlinear indicators must be freed and estimated unless the latent variables are uncorrelated. For example, consider the case where X in Eq. [3] is measured by three indicators, $x_1 -$

x_3 , and where Z is also measured by three indicators, $z_1 - z_3$, and where X and Z are correlated. In this case, the product-indicator pairs x_1z_1 , x_2z_2 , and x_3z_3 , could be used to measure the latent interaction term, with the product-indicator pairs x_1x_1 , x_2x_2 , and x_3x_3 , and z_1z_1 , z_2z_2 , and z_3z_3 , used to measure the two quadratic effects, respectively. In this case, the covariance between the errors of, for example, x_1z_1 and x_1x_1 , needs to be freely estimated, since both product-indicator terms share a common component in x_1 . We implemented this approach by adapting code from Kelava and Brandt (2009), with subsequent standardization of the coefficients following Brandt et al (2015) for comparability purposes. All data analysis was conducted with the *sem* package in R (Fox, 2016, version 3.1-8). Significance of the estimates was tested with the z tests, as was the case with LMS.

Results

An ideal analysis technique should both indicate an effect when such exists (no Type II errors) and not indicate a nonexistent effect (no Type I errors). Because these present two distinct scenarios, we present and discuss our results separately for the conditions with ($f^2 = 0.02, 0.15, \text{ or } 0.35$) and without ($f^2 = 0$) nonlinear effects, starting with the latter. Given our focus on the estimation of the nonlinear parts of model, we restrict our discussion of the results to only those for the three relevant paths (for the one interaction and two quadratic effects in Eq. [3])⁶.

Non-existing Nonlinear Effects – Estimate Accuracy

We start by assessing the absolute bias of the estimates in the scenario where all nonlinear effects were zero ($\gamma_3 = \gamma_4 = \gamma_5 = 0$ in Eq. [3]). These results are presented in Figure 1 (over levels of sample size), Figure 2 (over number of indicators), Figure 3 (over loading strength), and Figure 4 (over degree of correlation among the first-order latent variables), averaged over all other experimental factors. Throughout the figures bias was negligible, within +/- 0.005 range from zero for all techniques and experimental conditions. Therefore, none of the techniques produce any meaningful bias for nonlinear effects when such effects do not exist.

Insert Figures 1, 2, 3 and 4 About Here

Non-existing Nonlinear Effects – Type I Error (False Positive) Rate

While unbiasedness is certainly desirable, in most cases individual researchers have just a single estimate to interpret and therefore it is important to also understand how frequently the different techniques would lead to an incorrect inference. Therefore, we will next look at false positive rates, which for the $p < .05$ rule should not exceed 5% of the replications. As before, the results for false positive rates are presented over experimental factors in Figure 5 (sample size), Figure 6 (number of indicators), Figure 7 (loading strength), and Figure 8 (degree of correlation between the two first-order latent variables). The dashed line in the figures represents the 5% expectation for the Type I error rate and the techniques should ideally produce false positive rates that meet or at least do not greatly exceed this threshold.

Insert Figures 5, 6, 7 and 8 About Here

As before, on average, there are no major differences between the techniques. The overall Type I error rates for the interaction and quadratic estimates, respectively, were 5% for OLS and EXT, and 6% for DR, PLSc, and LMS. In contrast to the results on bias of the estimates where the performance of the estimators was very similar over the experimental factors, in the case of Type I error rate there were conditions with some noticeable deviations from these average rates. In particular, the Type I error rate of LMS was higher than its overall average for smaller samples and improving as sample size increases (see Figure 5) and also increased to about 7% when the indicator reliability was extremely high at 0.9. When considered over the number of indicators measuring each first-order latent variable, the Type I error rate for the EXT approach was lower (that is, more conservative) than the nominal rate when only 3 indicators were used (see Figure 6). All approaches except OLS showed an upward trend (that is, became more liberal than the nominal criterion) with increasing loading strength, which was most marked for the two latent variable approaches, LMS and EXT (see Figure 7). Finally, the degree of correlation between the two first-order latent variables did not have a noticeable effect for any of the statistical approaches, in all cases close to their overall rates for all levels of this simulation condition.

Taken together, the results for bias when estimating non-existing nonlinear effects indicate that the techniques had

remarkably similar performance. While the average 6% false positive rates for DR, PLSc, and LMS mean that these three approaches are slightly too liberal and violate the conservativeness principle in statistical analysis (cf., Aguirre-Urreta & Rönkkö, 2018), the violation is on average fairly small and therefore not a major concern for the applied IS researcher. However, we observed some conditions where there were some marked deviations from the nominal α level (see Figure 7). These findings should be carefully considered when making a choice as to statistical approach to be employed. We return to these in our discussion and recommendations for researchers.

Existing Nonlinear Effects – Estimate Bias

While there was no major difference between the techniques when all effects in the population were linear, the results show clear differences when nonlinear effects actually exist in the population; that is, when f^2 for the nonlinear effects is different from zero. The OLS estimates were severely biased across all simulation conditions due to the well-known attenuation effect. The magnitude of this bias was greater for the quadratic effects (in the -15% range) than for the interaction effect (in the -6% range). DR and PLSc produced nearly indistinguishable results that were mostly negatively biased, typically around -5% or less. LMS also produced generally negatively biased results, but the magnitude of the bias was smaller than with DR and PLSc. The EXT estimates, on the other hand, were generally positively biased (+5% or less).

Clear differences also emerge when inspecting the relative bias results by experimental condition. As before, we inspect the bias for the three paths one experimental factor at a time, averaging other factors. Figure 9 shows bias over sample size. The fact that all other estimators but OLS improve when sample size increases clearly illustrates the fact that whereas the other estimators are consistent under measurement error, OLS is not. Increasing the sample size reduced the bias of all consistent estimators (DR, PLSc, LMS, EXT). The effect is fairly strong up to $N = 300$ and levelling off afterwards. The results over effect size levels (i.e., f^2), shown in Figure 10, reveal that the performance of the estimators was relatively unaffected by the effect size.

Insert Figures 9 and 10 About Here

Figure 11 and Figure 12 present results over number of indicators and loading strength. Both these factors relate to measurement quality and have thus very similar effects on estimate bias. Starting with OLS, the estimates improve markedly as measurement quality increases. This result is expected because the attenuation effect depends on the reliability of the composites (sums) constructed from the raw items and the reliability of a composite depends on both the number of items as well as their reliabilities. The number of indicators or their loadings has a limited effect on estimates from DR, PLSc, and LMS, which are all nearly unbiased, though those from LMS less so across the board. Finally, there is a clear effect of both the number of indicators and their loadings on estimates obtained from the EXT approach. As before, these estimates are positively biased when compared to their true population values, and more so for the quadratic ones. In these two cases, the bias diminishes as measurement quality increases, but the effect is weaker for the interaction effect.

Insert Figures 11 and 12 About Here

Finally, Figure 13 shows the results by degree of correlation between the first-order latent variables in Eq. [3]. Again, the effects on LMS, DR and PLSc are small for both the interaction and quadratic effects. OLS and EXT estimates were affected, but only for the interaction and not the quadratic effect. This is a reasonable result, in that degree to which the first-order latent variables X and Z are correlated should not have any necessary effect on the quadratic estimates relating X^2 and Z^2 to the dependent variable. In the case of the interaction effect and OLS, the estimates are markedly biased downwards for low levels of this correlation but improve markedly (becoming slightly positively biased at the highest level of the correlation) as the degree to which the two first-order latent variables in Eq. [3] are correlated with each other. This highlights the fact that when multiple correlated predictors are contaminated with measurement error, it is difficult to say much about the magnitude or even the direction of bias (Wooldridge, 2009, p. 320) and the outcome can be surprising, as demonstrated by the recent study by Goodhue and colleagues (2017).

Insert Figure 13 About Here

These results clearly demonstrate the consequence of inconsistency of the OLS estimator, which performed the worst by a wide margin. In terms of bias of the estimate for an existing effect, the LMS method clearly outperformed other techniques being nearly universally least biased and the degree of bias was largely unaffected by the experimental factors, except sample size, which influenced all consistent estimators similarly. While the bias of LMS was often less than half of that of PLSc and DR, the performance of these estimators was also on adequate levels, at -5% or less and relatively unaffected by the experimental factors. With EXT, the performance varied somewhat over the experimental factors, and was often substantially worse and in a different direction than the other techniques.

Existing Nonlinear Effects – Statistical Power

Our final analysis focuses on statistical power. As before, these results are presented in a series of plots focusing on levels of one design factor and collapsed over the others in the simulation: sample size (Figure 14), effect size f^2 (Figure 15), number of indicators measuring each of the first-order latent variables in Eq. [3] (Figure 16), strength of the loadings relating these indicators to their latent variable (Figure 17), and degree of correlation between the two first-order predictors in Eq. [3] (Figure 18). For reference, the desirable level of 0.8 statistical power is shown with a dashed line in the upcoming plots.

Insert Figures 14, 15, 16, 17 and 18 About Here

Overall, the plots show that the ranking of the techniques is the same regardless of the experimental conditions; from stronger to weaker: OLS, LMS, EXT, and DR and PLSc (the latter two are undistinguishable from each other). The overall statistical power rates for the interaction and two quadratic paths are, respectively, 39%, 51%, and 50% for OLS, 34%, 48%, and 48% for LMS, 30%, 45%, and 45% for EXT, and 24%, 37%, and 37% for both DR and PLSc. At first sight, it may be surprising that OLS, which cannot account for the presence of measurement error in the individual indicators nor in the composites of those used as proxies for the latent variables of interest, turns out to be the most powerful approach. However, a similar pattern of results has also been observed in other studies of nonlinear modeling approaches (e.g., Cham, West, Ma, & Aiken, 2012). Referring to the LMS and EXT approaches, which are the only ones that operate directly on latent variables, Cham et al (2012) note that there is a tradeoff between accuracy of the estimates and the power to detect an effect; see also similar results from Ma, Aiken and West (2011) and Ledgerwood and Shrout (2011) for a similar discussion in the context of mediation processes. For the case of PLSc, the use of percentile confidence intervals with these techniques has been shown to be conservative in terms of Type I Error, but at the cost of overly wide – and therefore less powerful – confidence intervals (Aguirre-Urreta & Rönkkö, 2018).

Moving on to the effects of the individual design factors, the plots show general trends that are consistent with the known workings of statistical power. First, statistical power increases over sample size, so that larger samples have a higher likelihood of detecting an effect when said effect is present in the population. Second, statistical power also increases over effect size, so that larger effects have a higher likelihood of being detected than smaller effects. Third, statistical power increases as the number of indicators used to measure the latent variables increases. Given that, everything else being equal, having more indicators measuring a latent variable amounts to having more information about the latent variable, it is reasonable that statistical power increases with more information that can be used to assess the presence of an effect. Fourth, statistical power increases with stronger loadings relating the indicators to their latent variable. As before, this is a result of having more and better quality of measurement. Finally, statistical power decreases as the degree of the correlation between the first-order predictors increases because this lead also to higher correlation among the interaction and quadratic terms. This influences power not only by increasing multicollinearity of the model but also because, as a design feature of our simulation, the paths of individual terms decrease given that our simulation was parameterized to have a constant total effect size (f^2), which depends on both the correlation as well as the effects of the terms in the model.

If a researcher is simply interested in making a yes/no decision for the existence of a quadratic or interaction effect, the results for statistical power suggest that the OLS approach should be the preferred approach, particularly given that the false positive rate of this technique was at the nominal level of 5%. However, this recommendation is ill-advised for two reasons. First, based purely on statistical power considerations, the differences between OLS and LMS, which was the second-best performing technique in terms of statistical power, were not large. While LMS (as well as PLSc and DR) produces false positive rates that exceeded the nominal 5% level, this the overall rate of

false positives can still be deemed acceptable because this small difference could have occurred by chance only⁷. It should be noted, however, that the difference in power between OLS, with its biased estimates, and LMS, with its markedly more accurate results, is relatively small for the case of the interaction effect (an overall difference of 5%) and even more so for the quadratic effects (an overall difference of only 3%). Second, whereas in linear models the existence and direction of the effect can be inferred from p value and sign of the regression coefficient, this is not the case with nonlinear models, where the effect of an independent variable on the dependent variable is not constant but varies either as a function of itself (quadratic effect) or another variable (interaction effect). Therefore, the precision of the polynomial model estimates is not only important for judging the magnitude of an effect but also its shape and direction. To address this issue, we will next discuss the interpretation of polynomial models, demonstrating that the bias of OLS can lead to incorrect inference and, therefore, LMS or other consistent techniques should be strongly preferred over the inconsistent OLS.

To summarize, our simulation work shows the presence of important differences among the various techniques examined here with regards to our main outcomes of interest, namely estimate accuracy (or bias), type I error, and statistical power. These results are summarized in Tables 2 and 3. Specifically, Table 2 presents a summary of the results organized by technique (e.g., OLS, PLS, etc.), whereas Table 3 presents the same but organized by outcome (e.g., estimate accuracy, etc.).

Insert Tables 2 and 3 About Here

Marginal and Response Surface Plots and Interpretation of Polynomial Results

The interpretation of polynomial regression models is complicated by the fact that the magnitude of the coefficients not only defines the strength of the statistical association, but also its direction and shape. In fact, in a quadratic model the interpretation of the effect of X depends on the coefficients of the first order and quadratic term as well as the range of X . When the coefficient of X^2 is negative, the curve has an inverted U-shape and it would be tempting to interpret the effect as “first positive, then negative”. However, as demonstrated in the first plot of Figure 19, this interpretation is only warranted if the inflection point of the curve is within the range of X . For example, if we model “IS knowledge” as a quadratic function of age and estimate the first order effect at 3 and the quadratic term at -0.01, the inflection point would be at 150 years of age. Clearly, “first positive, then negative” interpretation would not be valid because no person has ever lived the 150 years required for the effect to turn negative. While this issue has received some attention in other disciplines (Haans, Pieters, & He, 2016), it has been largely ignored in IS, where researchers have without exception interpreted quadratic effects as evidence of a U-shape relationship ignoring the “diminishing returns” and “increasingly bad” interpretations shown in Figure 19.

With interactions of two variables, interpretation of the effects becomes even more complicated as those depend on the regression coefficients of both X and Z , and their product, as well as the range of both variables. The second plot in Figure 19 shows that, even if we hold the signs of the three regression coefficients constant, depending on their magnitudes and the range of both variables can mean that an effect of X is always positive, but varies in strength as a function of Z , can turn from positive to negative, or be always negative. Moreover, the effects can be such that if we divide the data in two groups according to Z , one group is always higher on Y than another, or it can be a cross-over effect so that which group is higher on Y depends on the value of X .

Insert Figure 19 About Here

As these examples demonstrate, the interpretation of polynomial regression effects is difficult and error-prone by simply looking at the raw regression estimates. Therefore, a number of tools including marginal effects, sometimes referred to as simple slopes analysis, and marginal predictions techniques and plots based on these calculations have been introduced to aid in interpretation of these models. Given that measurement error influences the magnitude of regression estimates, and the magnitude of the estimates together with the range of the variables define the correct interpretation, we will next introduce three different plots that are commonly used as interpretation aides – marginal effects plot, marginal prediction plot, and the response surface plot – and use these plots to demonstrate how measurement error can affect the interpretation of polynomial regression models. We note that these plots apply to results obtained from polynomial regression models irrespective of whether measurement error is present in the independent variables, thus making them useful to the universe of studies reviewed for Table 1

above.

Our example uses the well-known data for the European Consumer Satisfaction Index (ECSI) as adapted to the mobile phone market (Tenenhaus, Vinzi, Chatelin, & Lauro, 2005), which is distributed with the *semPLS* package in *R* (Monecke & Leisch, 2012) and is thus publicly available, allowing for full reproducibility of the example. We use Perceived Quality and Customer Expectations, as independent variables, and Perceived Value, as dependent variable, and estimate a polynomial model⁸ with one interaction and two quadratic effects. Perceived Quality is measured with 7 items, Customer Expectations with 3 items, and Perceived Value with 2 items, all on a 10-point scale. To demonstrate the effects of measurement error, we analyzed the data using OLS, where each construct was represented by the mean-centered unweighted average of its indicators, and by using LMS, where each construct was modeled as a latent variable, measured by its corresponding indicators. Results from these analyses are labeled 'Unstandardized' in Table 4. To make the two sets of estimates comparable, the regression coefficients should be expressed in a common metric. After scaling both analyses to a common metric, the OLS and LMS analyses are labeled 'Standardized' in Table 4.

Insert Table 4 About Here

As Table 4 shows, the estimates are quite different depending on whether measurement error is assumed to not exist (the OLS results) or when it is explicitly modeled by using a latent variable technique (the LMS results). While the estimates have always the same sign with both techniques, they differ in both magnitude and statistical significance. In this case, a researcher naively interpreting the results by looking at the significances would have reached different conclusions⁹. However, increasing the sample size would have likely made all estimates significant and thus the pattern or significances and signs of the estimates would have been identical between the two techniques. Yet, as we demonstrate next, the two sets of result shown in Table 4 lead to very different conclusions when interpreted with marginal effects plots, marginal prediction plots, or response surface plots.

As the name implies, marginal effect plots represent the marginal effects of the independent variables on the dependent variable at different levels of either the independent variable (quadratic effects) or a moderator variable (interaction effects). Marginal effect is the effect of increasing the value of one variable by a small amount, holding other variables at specific values. The effects are partial derivatives of the conditional mean function of the dependent variable with respect to each explanatory variable. Consider, for example, a linear model, such as the one presented in Eq. [4], containing two explanatory variables:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon \quad [4]$$

where β_0 is the intercept, β_1 and β_2 are the regression coefficients, X and Z are the independent variables, Y is the dependent variable, and ε is the disturbance term. The conditional mean function takes the following form (see Eq. [5]):

$$E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z \quad [5]$$

By partially differentiating the conditional mean function in Eq. [5] with respect to each of the independent variables, X and Z , we obtain their marginal effects: β_1 and β_2 , respectively. These are constants, as the model in Eq. [1] is linear-additive, and the effect of each independent variable is constant. Marginal effects become more interesting in polynomial models because they are no longer constant. In this case, the conditional mean function of Y with respect to X and Z takes the following form (see Eq. [6]):

$$E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_4 X^2 + \beta_5 Z^2 \quad [6]$$

Then, by partially differentiating the conditional mean of Y given X and Z with respect to each explanatory variable, we obtain the following two marginal effects (see Eq. [7] for the marginal effect of X and Eq. [8] for the marginal effect of Z):

$$\frac{\partial E(Y|X, Z)}{\partial X} = \beta_1 + \beta_3 Z + 2\beta_4 X \quad [7]$$

$$\frac{\partial E(Y|X, Z)}{\partial Z} = \beta_2 + \beta_3 X + 2\beta_5 Z \quad [8]$$

Therefore, when the polynomial model holds, the marginal effect of each independent variable is a function of both itself and the other independent variable in the model. Because the effect varies, we need to consider the different values that it can receive over the range of X and Z , which is precisely what marginal effects plots allow us to do. The advantage of marginal effects plot is that it allows us to study the magnitude of the effect as a function of two different variables at the same time. Figure 20 shows marginal effect plots for both OLS and LMS results. The plots are calculated by choosing one variable to be on the X axis (Customer Expectation) and plot a curve or a line demonstrating the marginal effect of the other variable (Z ; Perceived value). To assess how the marginal effect of the variable depends on itself, multiple lines can be added a few different values, typically at mean, and one standard deviation below, and one standard deviation above the mean.

Insert Figure 20 About Here

The two plots in Figure 20 are similar in shape, because they are both constructed by substituting the path estimates into the same formula. Nevertheless, there are important differences between the two. The first observation is that the two plots have very different scales on the y -axis. The reason for this is that the plots actually present two very different effects. Notably, the intercepts of the three slopes – the effects of Perceived Value when Customer Expectation is at zero – are quite different (much higher in the LMS plot than in the OLS plot). In addition, the slopes – changes in the effect of Perceived Value due to changes in Customer Expectations – are markedly different as well, in that Customer Expectations is clearly expected to have a much stronger impact on the effect of Perceived Value when the LMS results are considered. All in all, quite a different picture of the relationship between Customer Expectations and Perceived Value emerges when the results of LMS — which is an unbiased technique which can account for measurement error — are contrasted with those of OLS, which cannot do as much.

While a marginal effects plot is useful in that it allows interpreting the effect of one variable as a function of itself and another variable, the plot itself does not directly demonstrate the shape of the estimated relationship between the independent and dependent variable. This is done by marginal prediction plots. Marginal prediction plots are constructed by using the estimated path coefficients to calculate predicted values of the dependent variables across the range of one of the independent variables (plotted in the x -axis, in our example, Customer Expectations) and for fixed values of the other independent variable (plotted as separate curves, in our example, Perceived Quality). Instead of the size of the effect, the predicted values of the dependent variable (Perceived Value) are plotted on the y -axis. This plot type is demonstrated in Figure 21. As before, we chose three levels of Perceived Quality, at mean and ± 1 standard deviation from the mean. The plot presented in Panel (a) (OLS results) shows three curves, for the three different fixed values of Perceived Quality, that are rather flat and almost parallel, which would lead to the conclusion there is little to no nonlinearity present in the results. The plot presented in Panel (b) (LMS results), however, shows three markedly different curves, indicating not only that there is a nonlinear relationship (in this case, quadratic in shape) between Customer Expectations and Perceived Value, but also that the relationship is affected by the level of Perceived Quality at which it has been plotted.

Insert Figure 21 About Here

In general, we could conclude that Perceived Value increases with increasing Customer Expectations up to a point, above which increasing Customer Expectations are associated with decreasing Perceived Value. Moreover, both the level of Customer Expectations at which that maximum occurs, as well as the overall level of Perceived Value reached at that point, are dependent on the value of Perceived Quality, so that when the value of Perceived Quality is low (e.g., -1 SD) the maxima of Perceived Value is both lower and occurs at a lower value of Customer Expectations compared to higher values of Perceived Quality.

Finally, we turn to response surface analysis as a means for better interpretation of the results of polynomial regression models. Response surface plots are similar to marginal prediction plots in that they are constructed by using path estimates to obtain predicted values of the dependent variable for various combinations of levels of the two explanatory ones. These plots are, however, three-dimensional in nature, and allow for a more comprehensive examination of the polynomial model of interest. See Figure 22 for response surface plots created with the estimates obtained from the OLS and LMS results.

Insert Figure 22 About Here

The two plots are strikingly different. The plot shown in Panel a (OLS results) presents a response surface that is almost flat in shape (e.g., a plane), giving little indication that there are any nonlinear effects present. The plot shown in Panel b (LMS results), on the other hand, provides clear evidence for the presence of a strong nonlinear relationship between the two predictor variables in how they affect the dependent one. In particular, it is clear that the effects of both explanatory variables are heavily dependent, and in a nonlinear way, on the particular values of the other explanatory variable. Even without engaging in a full response surface analysis, it is evident that ignoring the effects of measurement error in the obtained estimates leads to very different interpretations of the results.

Discussion and Recommendations

The use of polynomial regression models has become quite popular in the IS discipline as of late and the discipline is also increasingly aware of the need to correct estimation results for unreliability (Dijkstra & Henseler, 2015). However, the negative effects of measurement error on polynomial regression model estimates are not widely recognized. Therefore, the first goal of our study was to compare the performance of various statistical techniques that can be used to estimate polynomial regression models and correct for measurement error. To further encourage the adoption of these practices in the discipline, we also included the results from OLS regression to serve as a benchmark and to demonstrate what the effects of ignoring measurement error can be.

Our results show that, when the first-order predictors are measured with error, the estimates of the effects of the first-order, multiplicative, and higher-order (e.g., powered) terms are all markedly biased. Though this is a well-known consequence of estimating these models with OLS regression, which assumes no measurement error in the independent variables, we believe the fact that the effect is much more severe in product and higher-order terms has not been sufficiently recognized in the IS discipline. Therefore, we strongly advise researchers against the continued use of OLS regression, which is currently the most common technique for the estimation of these models, unless the level of measurement error can be demonstrated to be negligible. Furthermore, we also considered two different families of estimators that take into consideration the presence of measurement error in the predictor variables. The first of these, which includes Consistent PLS (PLSc) and Disattenuated Regression (DR), does so by adjusting estimates from an initial run through disattenuation formulas. The second group of estimators, which includes the Latent Moderated Squares (LMS) and Extended Unconstrained (EXT) approaches, work directly with latent variables and thus incorporate the quality of measurement (e.g., reliability at the individual indicator level) directly into the estimation of the parameters.

Our results indicate that, in the absence of a nonlinear effect, all techniques considered in the study perform well (in terms of deviations from zero estimates and Type I error rates). When there is a nonlinear effect, however, LMS tends to outperform all other techniques. It is not only an essentially unbiased estimator of the relationships of interest, but it is also the most powerful — in terms of ability to detect the presence of an effect when an effect is indeed present in the population; e.g., statistical power — than all the other approaches. The other technique based on latent variables, EXT, also performs well but tends to exhibit less statistical power. The two techniques based on corrections for the presence of measurement error, PLSc and DR, tend to under-correct, thus providing estimates that are biased downwards, as well as are less powerful than the two approaches based on latent variables. Given the relative ease of use of LMS and the existence of full software support for it (e.g., MPlus and the nlsem package in R), it would be natural to recommend this approach as the default way for estimating of polynomial regression models.

However, there is an important limitation shared by LMS, PLSc, and DR in that all these techniques assume that the construct of interest is normally distributed, and LMS further assumes normally distributed errors. If these assumptions do not hold, all these techniques can produce misleading results. Fortunately, neither of these normality assumptions is made by the EXT approach, which thus provides a safer alternative. It is therefore important to justify particularly the normality assumption about the latent variable when applying these techniques. Such justification can come from theoretical arguments, empirical tests, or ideally both. While the distribution of a latent variable cannot be estimated directly, whether the normality assumption can be expected to hold can be done indirectly by comparing the LMS estimates against the EXT estimates using the Hausman specification test (Greene, 2012, sec. 8.4.1). A thorough explanation of the test is beyond the scope of this research, but the objective of the test is to calculate whether the difference between a consistent (EXT) and a potentially inconsistent (LMS) estimate can be attributed to chance only by comparing the difference against its standard errors, similarly to z or t statistics for tests of one estimate, or a χ^2 test for multiple estimates. While the same technique could in principle be applied to compare also EXT against PLSc and DR, this comparison would require rescaling the estimates to a common

metric, which may be cumbersome.

The second main goal of our research was to aid in the interpretation, for both authors and readers alike, of results obtained from the estimation of polynomial regression models. We do so by presenting three distinct plots, and their interpretation, that can be used to illustrate nonlinear effects, as well as illustrate them through a worked example from a well-known dataset. In particular, we considered marginal effect plots, marginal prediction plots, and response surface plots, all of which can be constructed from the results of a polynomial regression model (we recommend that the estimates used to construct the plot be obtained from a technique shown to provide accurate results, such as LMS, per our earlier discussion). For each of these plots we discussed their construction as well as interpretation. Given that inherent challenges in understanding nonlinear effects, we believe the inclusion of a graphical representation of those would greatly aid in their interpretation, and thus recommend researchers incorporate those into future publications.

Standardization and centering, while very common in IS, can be problematic when applying these techniques. We note that standardized regression is particularly problematic because this involves also standardization of the product of the indicators, thus making the estimates of the interaction biased and inconsistent (Dawson, 2014). If standardized results are desired, they should be calculated by standardizing the variables instead of the regression results. However, both standardization and centering make interpreting the plots more difficult because the plots will be on the standardized instead of the original metric. This is important not only when the variable has a natural scale (e.g. a binary yes/no variable, euros, years, and so on), but also when it is measured on a rating scale. Keeping the variables on their original scale (e.g. by taking a mean of the scale) thus allows to compare the effect between informants based on their average scale response (e.g. 2, 4, and 6 on a 7-point scale) instead of means and standard deviations that are sample specific quantities and can vary from one study to another and are less transparent. In fact, the use of the original scale of the indicators is exactly what LMS and EXT do.

Our research provides clear evidence of the negative consequences of the common practice of ignoring measurement error in the independent variables in polynomial regression models. Fortunately, the issue can be addressed by using several statistical techniques that are also increasingly available as parts of standard software packages. Going forward, we hope that our study not only raises awareness of the issue of measurement error in polynomial models, but also steers researchers toward more rigorous analysis using modern techniques as well as more rigorous interpretation of the results using any of the graphical techniques that were presented in this research.

References

- Aguirre-Urreta, M., Marakas, G. M., & Ellis, M. E. (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.
- Aguirre-Urreta, M., & Rönkkö, M. (2018). Statistical inference with PLSc using bootstrap confidence intervals. *MIS Quarterly*, 42(3), 1001–1020.
- Brandt, H., Umbach, N., & Kelava, A. (2015). The Standardization of Linear and Nonlinear Effects in Direct and Indirect Applications of Structural Equation Mixture Models for Normal and Nonnormal Data. *Frontiers in Psychology*, 6, 1–11.
- Bring, J. (1994). How to Standardize Regression Coefficients. *The American Statistician*, 48(3), 209–213.
- Cham, H., West, S., Ma, Y., & Aiken, L. (2012). Estimating Latent Variable Interactions with Non-Normal Observed Data: A Comparison of Four Approaches. *Multivariate Behavioral Research*, 47(6), 840–876.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods*, 2(2), 161–172.
- Cortina, J. M., & Landis, R. S. (2011). The earth is not round ($p = .00$). *Organizational Research Methods*, 14(2), 332–349.
- Culpepper, S. A. (2012). Evaluating EIV, OLS, and SEM estimators of group slope differences in the presence of measurement error the single-indicator case. *Applied Psychological Measurement*, 36(5), 349–374.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29.
- Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, 15(3), 339–362.
- Dawson, J. F. (2014). Moderation in Management Research: What, Why, When, and How. *Journal of Business and Psychology*, 29(1), 1–19.
- Devlieger, I., & Rosseel, Y. (2017). Factor Score Path Analysis. *Methodology*, 13(Supplement 1), 31–38.

- Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297–316.
- Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrical*, 79(4), 585–604.
- Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology*, 3(3), 100–114.
- Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). San Francisco, CA: Jossey-Bass.
- Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, 36(6), 1577–1613.
- Fox, J. (2016). *sem: Structural Equation Models*.
- Fuller, W. A. (1987). *Measurement error models*. New York, NY: Wiley.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2017). A Multicollinearity and Measurement Error Statistical Blind Spot: Correcting for Excessive False Positives in Regression and PLS. *MIS Quarterly*, 41(3), 667-A15.
- Greene, W. H. (2012). *Econometric analysis*. Boston: Prentice Hall.
- Haans, R. F. J., Pieters, C., & He, Z.-L. (2016). Thinking about U: Theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strategic Management Journal*, 37(7), 1177–1195.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: a global perspective*. (7th ed.). Upper Saddle River, NJ: Pearson Education.
- Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, 117(2), 348–357.
- Jöreskog, K., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. Marcoulides & R. Schumaker (Eds.), *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, N.J: Lawrence Erlbaum Associates.
- Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, 16(2), 123–131.
- King, G. (1986). How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science*, 30(3), 666–687.
- Klein, A., & Moosbrugger, H. (2000). Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method. *Psychometrika*, 65(4), 457–474.
- Lankton, N. K., McKnight, D. H., Wright, R. T., & Thatcher, J. B. (2016). Research Note—Using Expectation Disconfirmation Theory and Polynomial Modeling to Understand Trust in Technology. *Information Systems Research*, 27(1), 197–213.
- Ledgerwood, A., & Shrout, P. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, 101(6), 1174–1188.
- Ma, Y., Aiken, L., & West, S. (2011). Comparison of statistical power for testing interactive versus quadratic effects using measured variable versus latent variable approaches. Paper presented at the meeting of the American Statistical Association. Miami, FL.
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *MIS Quarterly*, 35(2), 293–334.
- Marsh, H., Wen, Z., & Hau, K. (2004). Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction. *Psychological Methods*, 9(3), 275–300.
- Monecke, A., & Leisch, F. (2012). semPLS: Structural Equation Modeling Using Partial Least Squares. *Journal of Statistical Software*, 48(3), 1–32.
- Moody, G., Lowry, P., & Galletta, D. (2017). It's Complicated: Explaining the Relationship between Trust, Distrust, and Ambivalence in Online Transaction Relationships using Polynomial Regression Analysis and Response Surface Analysis. *European Journal of Information Systems*, 26(4), 379–413.
- Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches. *Structural Equation Modeling in Educational Research: Concepts and Applications*, 103–136.
- Muthén, B., & Muthén, L. (2004). *MPlus 3.0 User's Guide*. Los Angeles, CA: Author.
- Myers, R., & Anderson-Cook, C. (2016). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Hoboken, NJ: John Wiley & Sons, Inc.
- Ping, R. (1996). Latent Variable Regression: A Technique for Estimating Interaction and Quadratic Coefficients. *Multivariate Behavioral Research*, 31(1), 95–120.

- R Core Team. (2016). *R: A Language and Environment for Statistical Computing* (Version 3.3.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Ree, M., & Carretta, T. (2006). The Role of Measurement Error in Familiar Statistics. *Organizational Research Methods*, 9(1), 99–112.
- Rönkkö, M. (2017). *matrixpls: Matrix-based partial least squares estimation* (Version 1.0.5). Retrieved from <https://github.com/mronkko/matrixpls>
- Rönkkö, M., McIntosh, C. N., & Aguirre-Urreta, M. I. (2016). Improvements to PLSc: Remaining problems and simple solutions. Unpublished Working Paper. Retrieved from <http://urn.fi/URN:NBN:fi:aalto-201603051463>
- Shanock, L., Baran, B., Gentry, W., Pattison, S., & Heggestad, E. (2010). Polynomial Regression with Response Surface Analysis: A Powerful Approach for Examining Moderation and Overcoming Limitations of Difference Scores. *Journal of Business and Psychology*, 25, 543–554.
- Singh, H., Aggarwal, R., & Cojuharenco, I. (2015). Strike a Happy Medium: The Effect of IT Knowledge on Venture Capitalists' Overconfidence in Investments. *MIS Quarterly*, 39(4), 887–908.
- Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of AIS*, 2004(13), 380–427.
- Tenenhaus, M., Vinzi, V., Chatelin, Y., & Lauro, C. (2005). PLS Path Modeling. *Computational Statistics & Data Analysis*, 48, 159–205.
- Umbach, N., Naumann, K., Hoppe, D., Brandt, H., Kelava, A., & Schmitz, B. (2015). *nlsem: Fitting structural equation mixture models* (Version 0.3). Retrieved from <http://cran.r-project.org/web/packages/nlsem/index.html>
- Venkatesh, V., & Goyal, S. (2010). Expectation disconfirmation and technology adoption: polynomial modeling and response surface analysis. *MIS Quarterly*, 34(2), 281–303.
- Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed). Mason, OH: South Western, Cengage Learning.
- Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed). Mason, OH: South-Western Cengage Learning.

About the Authors

Miguel I. Aguirre-Urreta is a visiting associate professor in the Department of Information Systems and Business Analytics, College of Business, at Florida International University. Before joining FIU, he was on the faculty at Texas Tech University and DePaul University. Miguel received his Ph.D. 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 degree is in public accounting from the Universidad de Buenos Aires, in Argentina. Before entering academia, he was an auditor with Arthur Andersen and an external reporting analyst with Movicom Bellsouth. Miguel is interested in quantitative research methods, computer self-efficacy, human–computer interaction, technology acceptance and diffusion, and data science. He is also a past chair of the *AIS Special Interest Group on Human-Computer Interaction*. His research has appeared or is forthcoming at *MIS Quarterly*, *Information Systems Research*, *Communications of the AIS*, *Research Synthesis Methods*, *IEEE Software*, *Measurement*, and *The DATA BASE for Advances in Information Systems*, among others.

Mikko Rönkkö is postdoctoral researcher at University of Jyväskylä. He completed his Ph.D about misuse of structural equation modeling and other statistical techniques in management research at Aalto University in 2014. His current research interests are statistical research methods and software entrepreneurship. He has taught courses on statistical research methods in multiple universities in and outside of Finland. He has published in *Organizational Research Methods*, *Psychological Methods*, *Journal of Operations Management*, *MIS Quarterly*, among other journals. He is a department editor in *Journal of Operations Management* and on the editorial boards of *Organizational Research Methods* and *Entrepreneurship Theory and Practice*.

Jiang Hu is a Ph.D student at College of Business, University of Texas at Arlington. He received his master's degree in information systems and supply chain management from Naveen Jindal School of Management at the University of Texas at Dallas. His research interests include information privacy, sharing economy and healthcare IT. His research has been presented at the Americas Conference on Information Systems.

Table 1. Summary of Research (2012-2016)

Journal	Measurement Error in IVs	Analytical Technique	Number of Articles
MIS Quarterly	Yes	OLS	4
	No	OLS, Tobit	6
Information Systems Research	Yes	OLS	3
	No	OLS, Cobb-Douglas, Hausman-Taylor	11
Journal of the Association for Information Systems	Yes	OLS	1
	No	---	---
Journal of Management Information Systems	Yes	OLS (3), PLS (1)	4
	No	OLS, Random Effects, Hausman-Taylor	5

Table 2. Summary of Results by Technique

Ordinary Least Squares (OLS)
<ul style="list-style-type: none"> – Limited (in the negligible range) estimate bias when there are no non-linear effects in the population; no marked effect of simulation factors on this outcome – Type I error in the nominal range (5%); no marked effect of simulation factors on this outcome – Most biased estimates of all techniques when there are non-linear effects in the population, in the -7% range for the interaction path and -15% range for the quadratic paths; marked effect of loading strength and number of indicators, also marked effect of first-order predictor correlation (but only for the interaction path) – Most powerful technique of those examined; statistical power varies as expected based on statistical theory
Consistent Partial Least Squares (PLSc)
<ul style="list-style-type: none"> – Limited (in the negligible range) estimate bias when there are no non-linear effects in the population; no marked effect of simulation factors on this outcome – Type I error above the nominal value (6%); no marked effect of simulation factors on this outcome – Negatively-biased estimates when there are non-linear effects in the population and undistinguishable from those of DR; no marked effect of simulation factors on this outcome – Least powerful technique of those examined (and undistinguishable from DR); statistical power varies as expected based on statistical theory
Disattenuated Regression (DR)
<ul style="list-style-type: none"> – Limited (in the negligible range) estimate bias when there are no non-linear effects in the population; no marked effect of simulation factors on this outcome – Type I error above the nominal value (6%); no marked effect of simulation factors on this outcome – Negatively-biased estimates when there are non-linear effects in the population and undistinguishable from those of PLSc; no marked effect of simulation factors on this outcome – Least powerful technique of those examined (and undistinguishable from PLSc); statistical power varies as expected based on statistical theory
Latent Moderated Squares (LMS)
<ul style="list-style-type: none"> – Limited (in the negligible range) estimate bias when there are no non-linear effects in the population; no marked effect of simulation factors on this outcome – Type I error above the nominal value (6%); some deviations from the average for different levels of sample size and loading strength – Negatively-biased and least biased estimates of all techniques when there are non-linear effects in the population; no marked effect of simulation factors on this outcome – Second most powerful technique of those examined (behind OLS, but differences are small in most conditions); statistical power varies as expected based on statistical theory
Extended Unconstrained Approach (EXT)
<ul style="list-style-type: none"> – Limited (in the negligible range) estimate bias when there are no non-linear effects in the population; some differences (small in range) across levels of each simulation factor – Type I error in the nominal range (5%); some variation between weaker and stronger loading conditions – Positively-biased estimates when there are non-linear effects in the population; some deviations from average (but always positively-biased) depending on sample size, number of indicators, and loading strength – Third most powerful technique of those examined (behind OLS and LMS); statistical power varies as expected based on statistical theory

Table 3. Summary of Results by Outcome

Non-existing Nonlinear Effects – Estimate Accuracy
<ul style="list-style-type: none"> – Bias was negligible, within +/- 0.005 range from zero for all techniques and experimental conditions – None of the techniques produce any meaningful bias for nonlinear effects when such effects do not exist
Non-existing Nonlinear Effects – Type I Error (False Positive) Rate
<ul style="list-style-type: none"> – No major differences between techniques – Type I error rates were 5% for OLS and EXT, and 6% for DR, PLSc, and LMS – LMS was higher than its overall average for smaller samples and improving as sample size increases; also increased to about 7% when the indicator reliability was extremely high at 0.9 – EXT approach was lower (that is, more conservative) than the nominal rate when only 3 indicators were used – All approaches except OLS showed an upward trend with increasing loading strength; most marked for the two latent variable approaches, LMS and EXT
Existing Nonlinear Effects – Estimate Bias
<ul style="list-style-type: none"> – Techniques ranked from most to least biased: OLS (by a wide margin), DR and PLSc, LMS (all four negatively biased), and EXT (positively biased) – Estimate accuracy improves with sample size, except for OLS – Effect size has no noticeable effects on estimate accuracy – OLS and EXT estimates improve (become more accurate) as loading strength and number of indicators increase; limited effects on DR, PLSc, and LMS – Correlation between first-order predictors only affects OLS and EXT estimates, and only those for the interaction path
Existing Nonlinear Effects – Statistical Power
<ul style="list-style-type: none"> – Techniques ranked from stronger to weaker: OLS, LMS, EXT, and DR and PLSc (the latter two are undistinguishable) – Caveat with regards to OLS: not much stronger than LMS, but at the cost of markedly more biased estimates – Statistical power increases over sample size, effect size, number of indicators, and loading strength – Statistical power decreases as the degree of the correlation between the first-order predictors increases – The ideal of 80% statistical power not reached in our simulation design, which was extensive in coverage

Table 4. Results from the ECSI Dataset

Coefficient	OLS			LMS		
	Unstand.	Standard.	Sig.	Unstand.	Standard.	Sig.
CUEX	0.100	0.127	ns	0.338	0.198	ns
PERQ	0.774	0.980	***	1.147	0.510	ns
CUEX x PERQ	0.059	0.095	ns	- 4.365	2.741	***
CUEX ²	- 0.027	- 0.043	ns	- 2.533	- 1.391	***
PERQ ²	- 0.012	- 0.018	ns	- 1.863	- 1.337	***

Note: CUEX = Customer Expectations, PERQ = Perceived Quality, *** $p < 0.001$

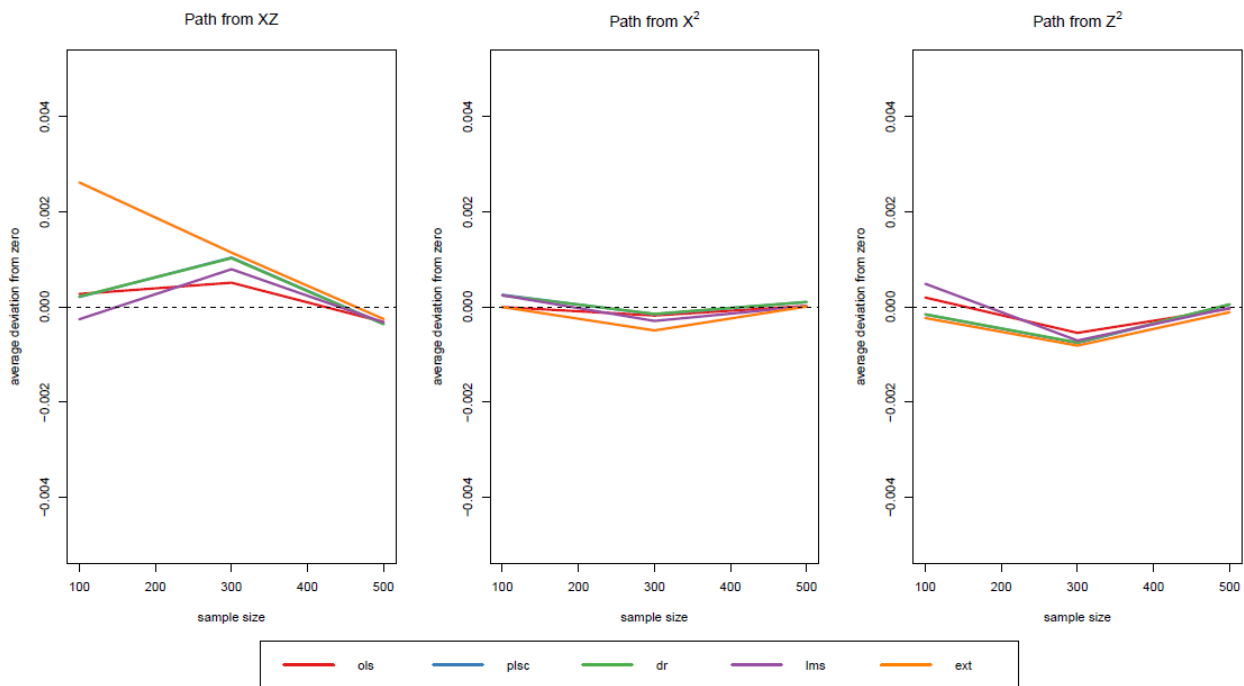


Figure 1. Average Deviation from Zero over Sample Size

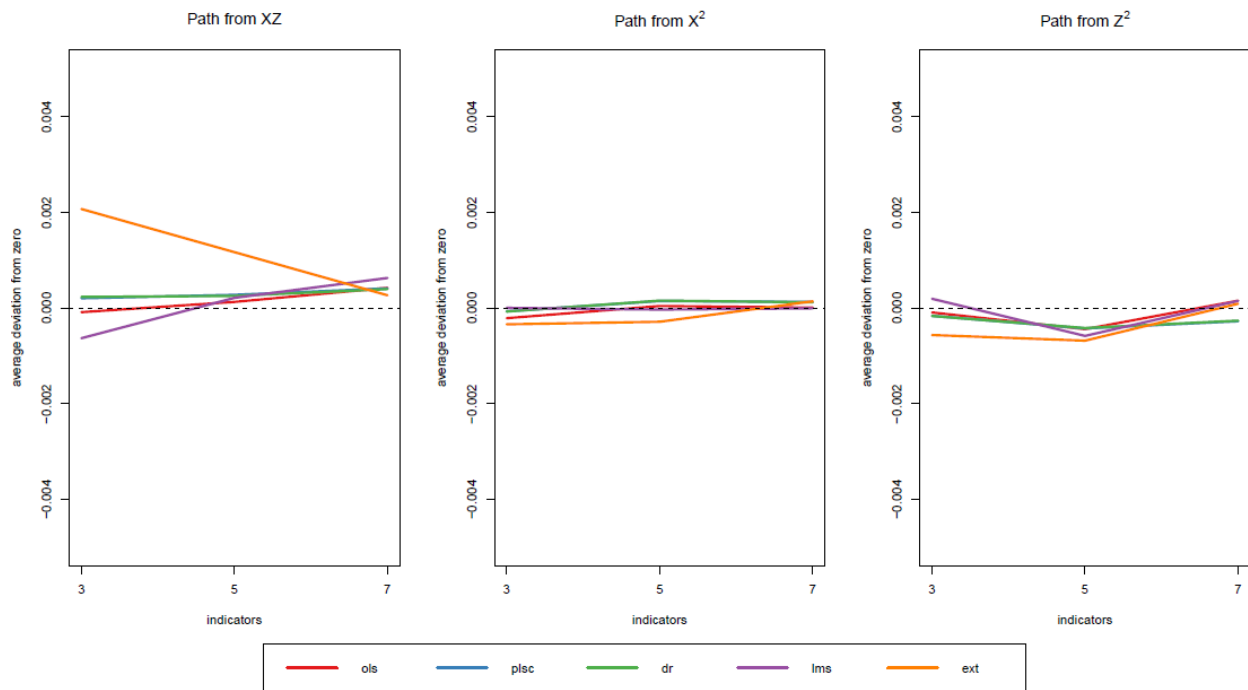


Figure 2. Average Deviation from Zero over Number of Indicators

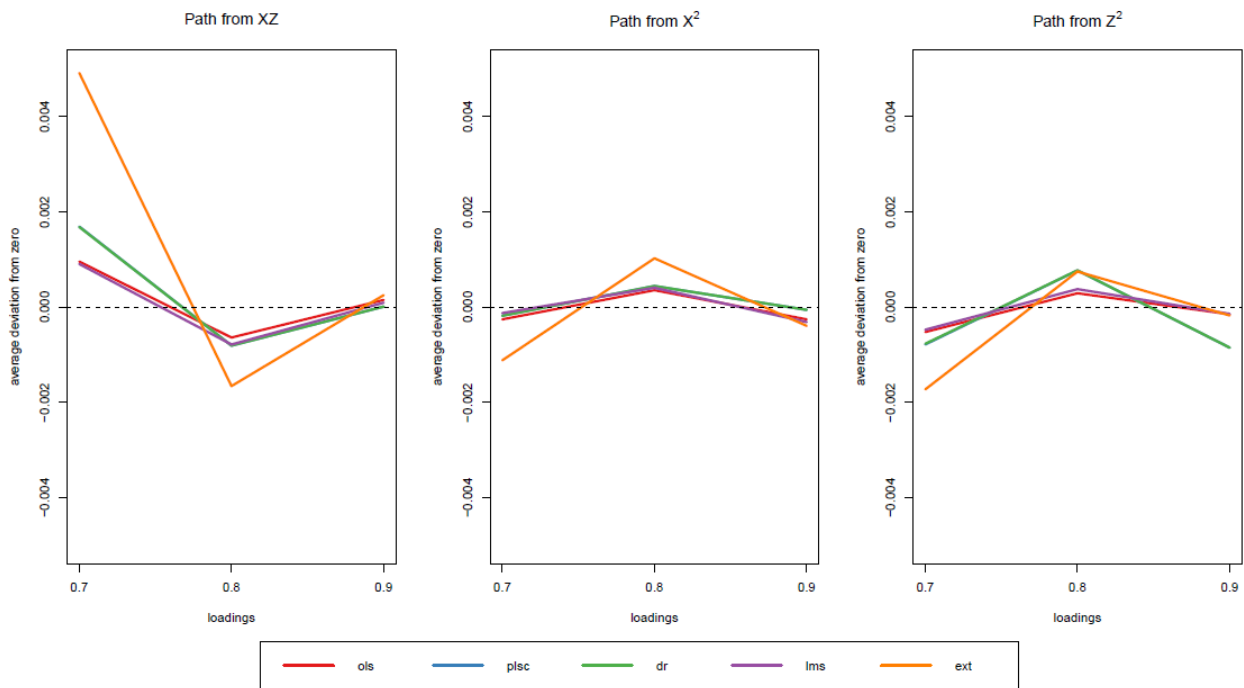


Figure 3. Average Deviation from Zero over Loading Strength

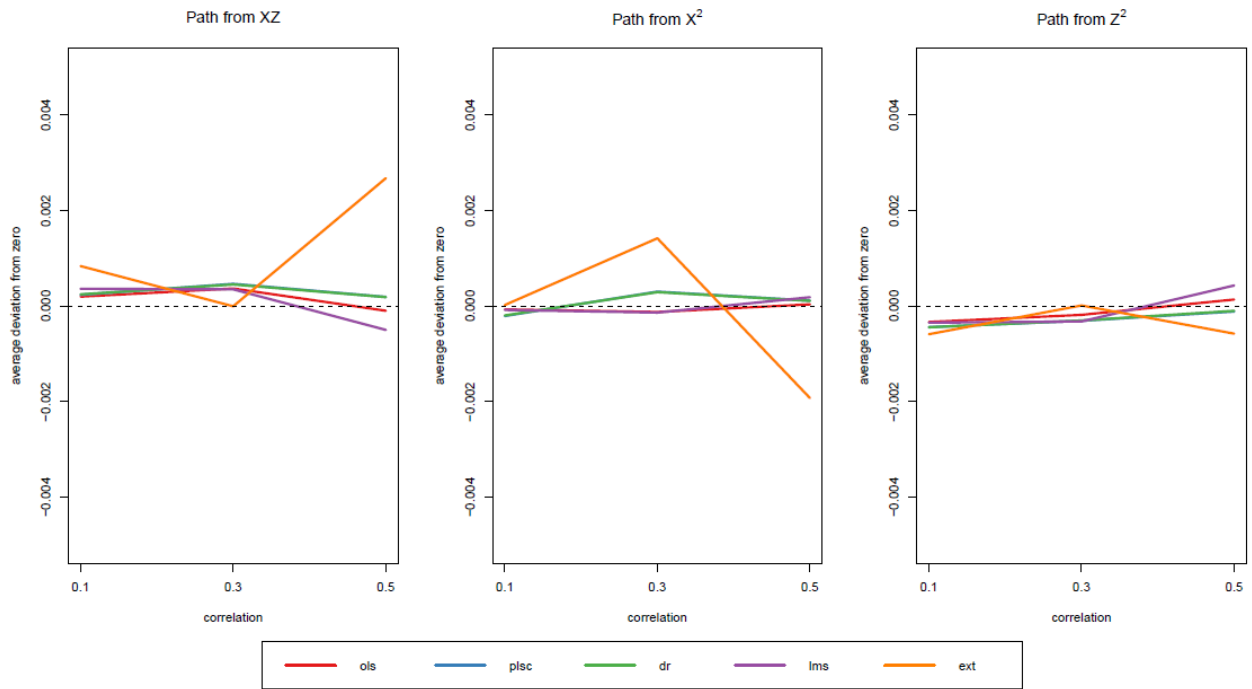


Figure 4. Average Deviation from Zero over Latent Variable Correlation

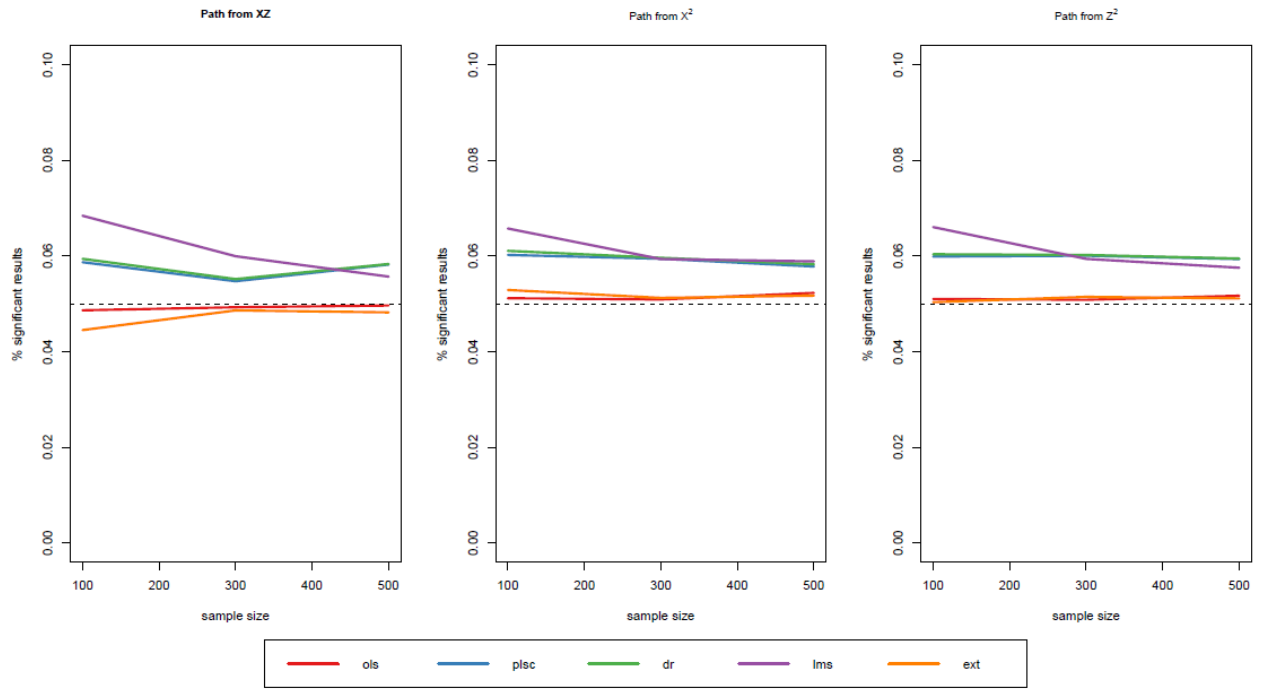


Figure 5. Type I Error over Sample Size

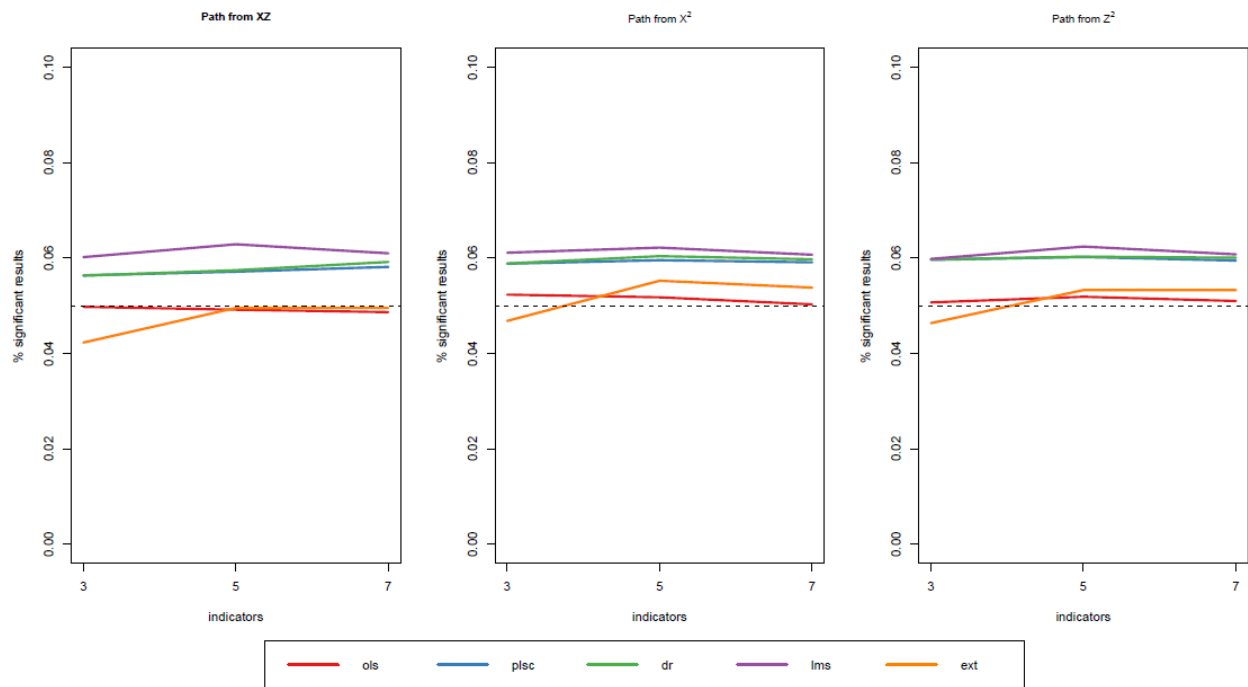


Figure 6. Type I Error over Number of Indicators

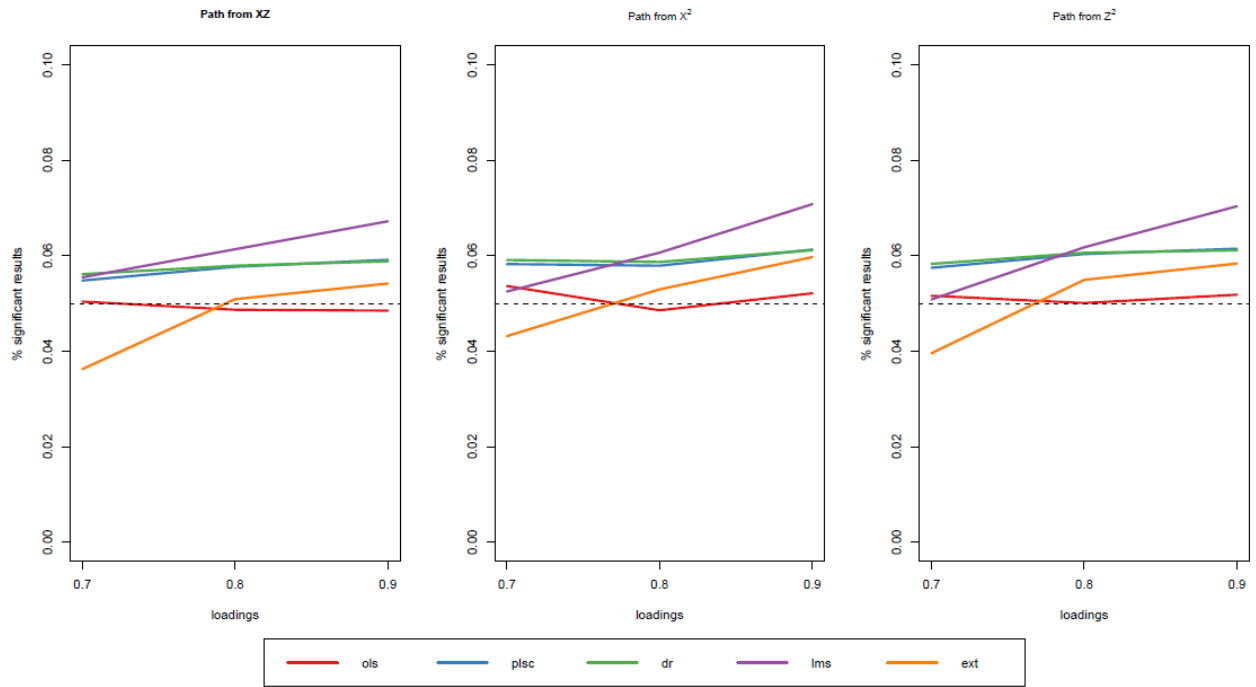


Figure 7. Type I Error over Loading Strength

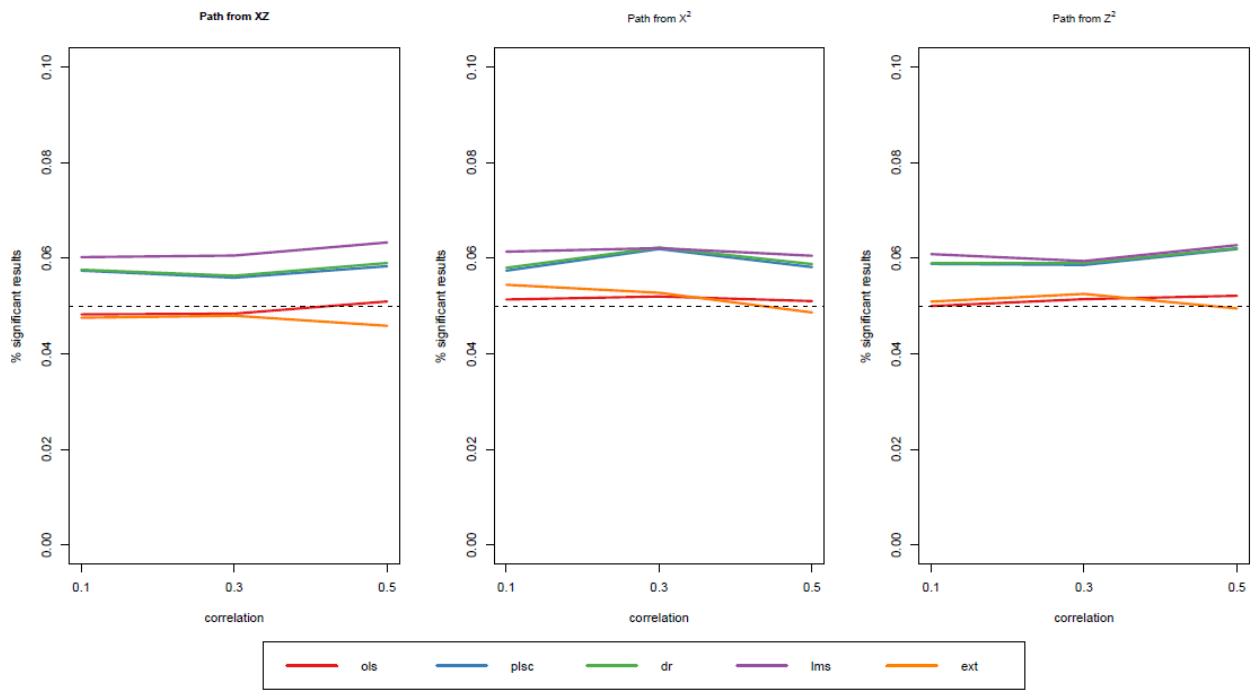


Figure 8. Type I Error over Latent Variable Correlation

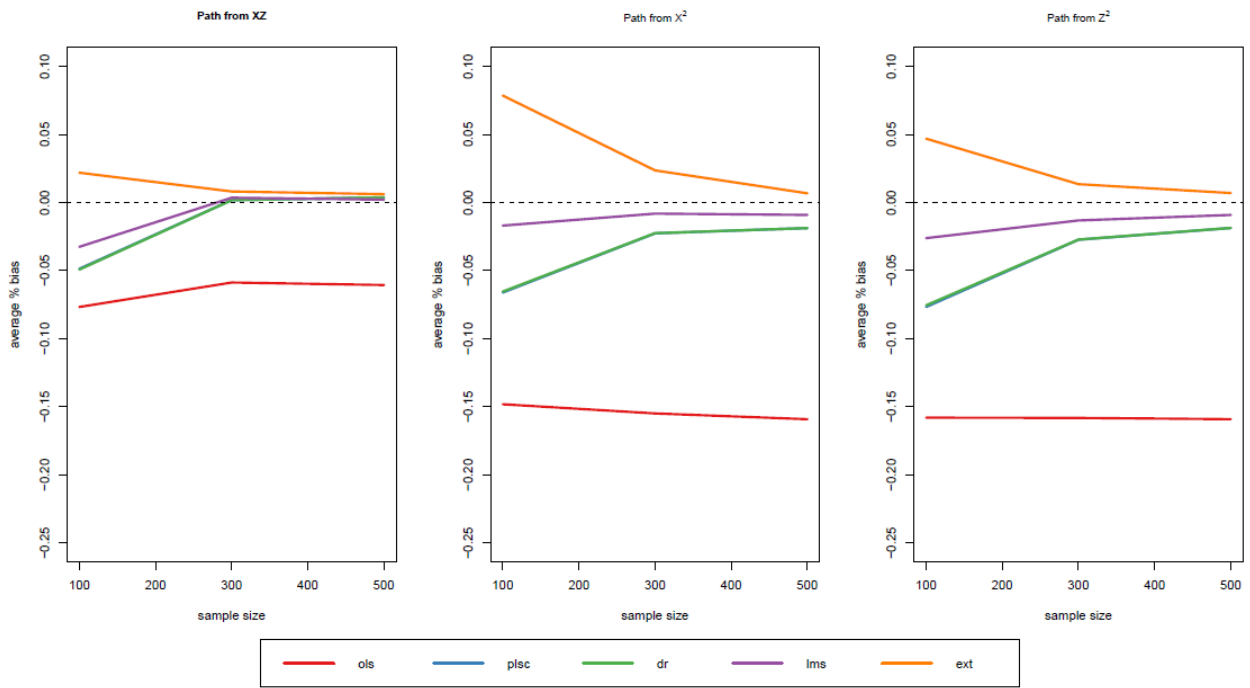


Figure 9. Average Estimate Bias (as % of Population Value) over Sample Size

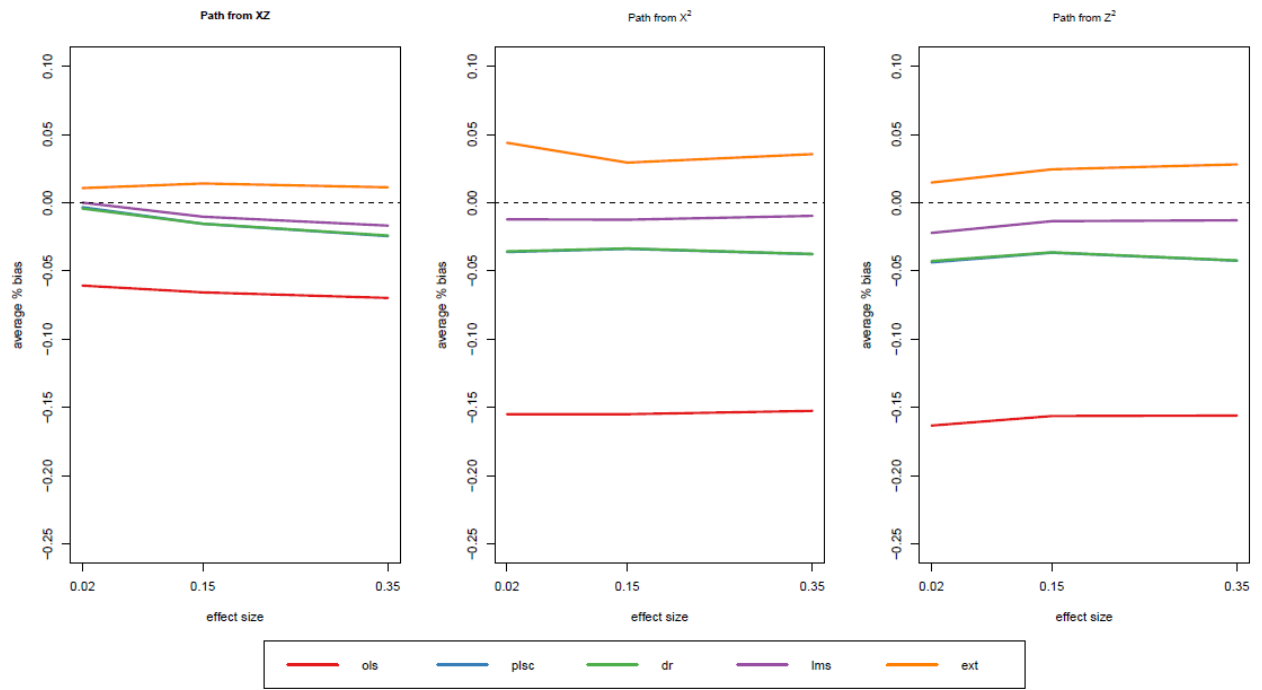


Figure 10. Average Estimate Bias (as % of Population Value) over Effect Size (f^2)

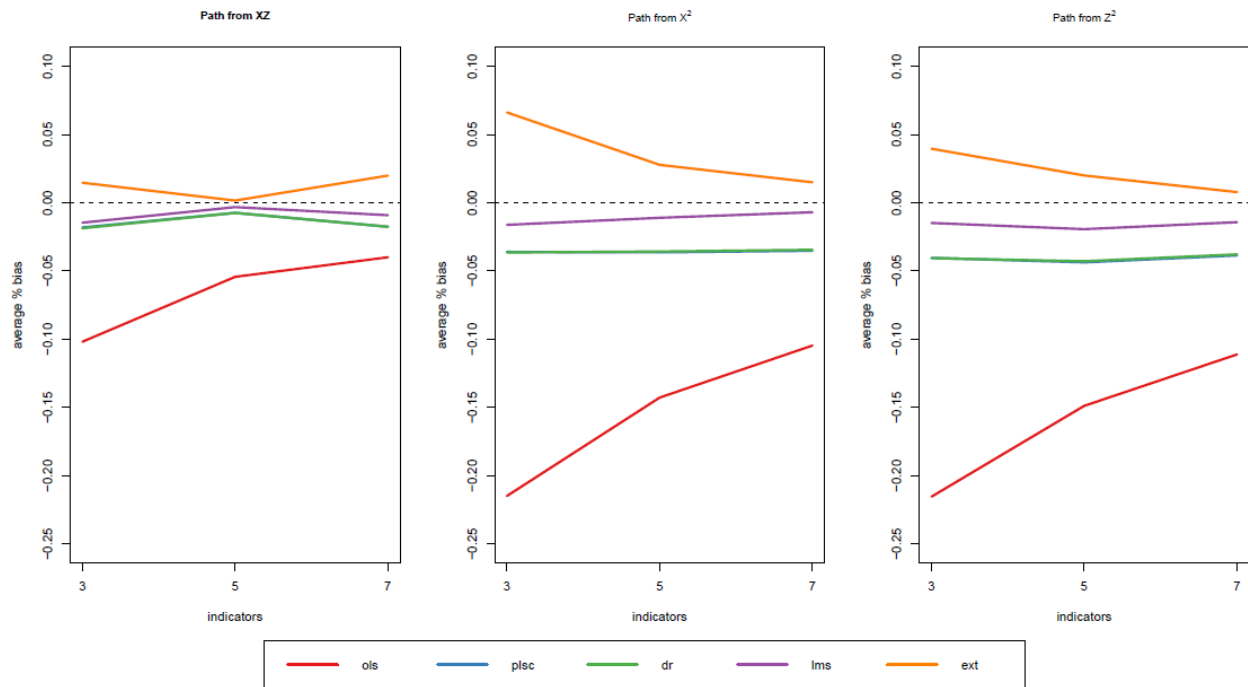


Figure 11. Average Estimate Bias (as % of Population Value) over Number of Indicators

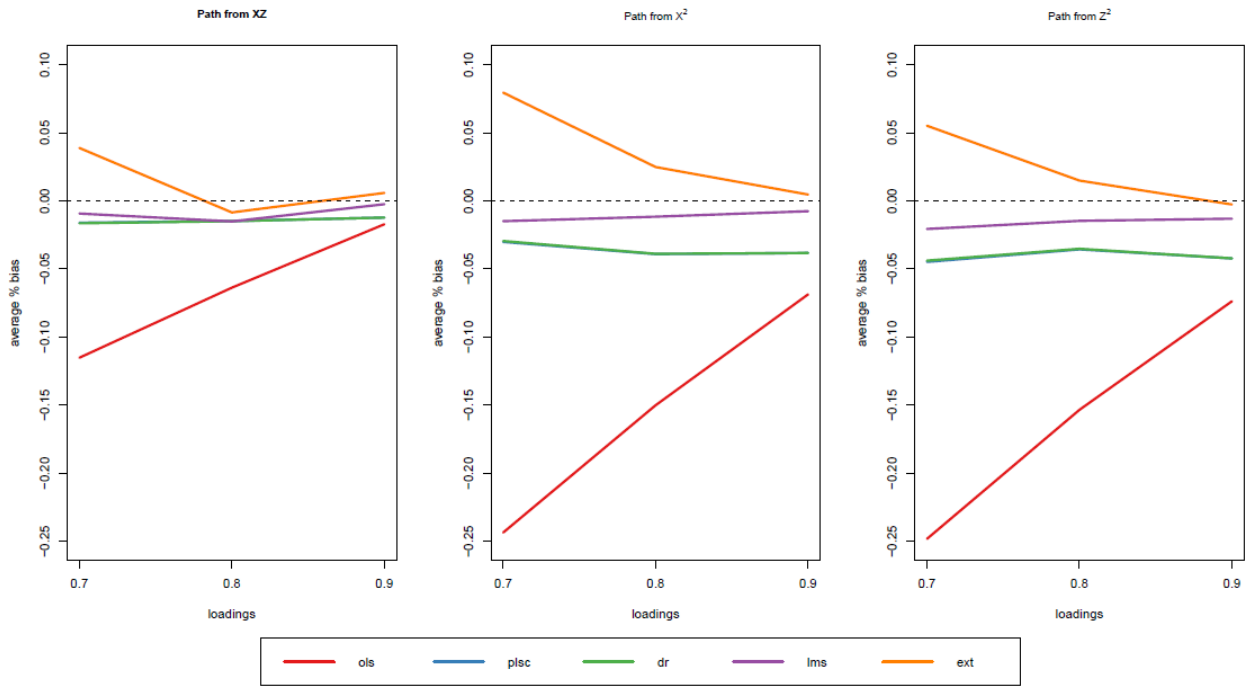


Figure 12. Average Estimate Bias (as % of Population Value) over Loading Strength

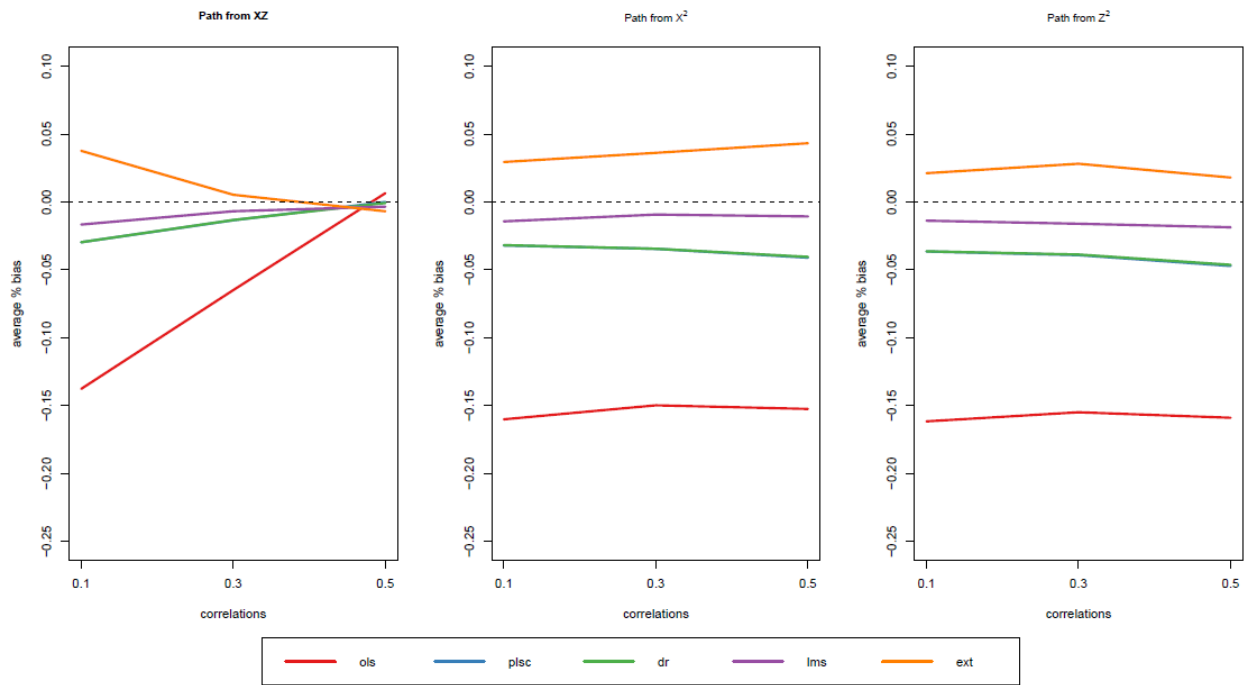


Figure 13. Average Estimate Bias (as % of Population Value) over Latent Variable Correlation

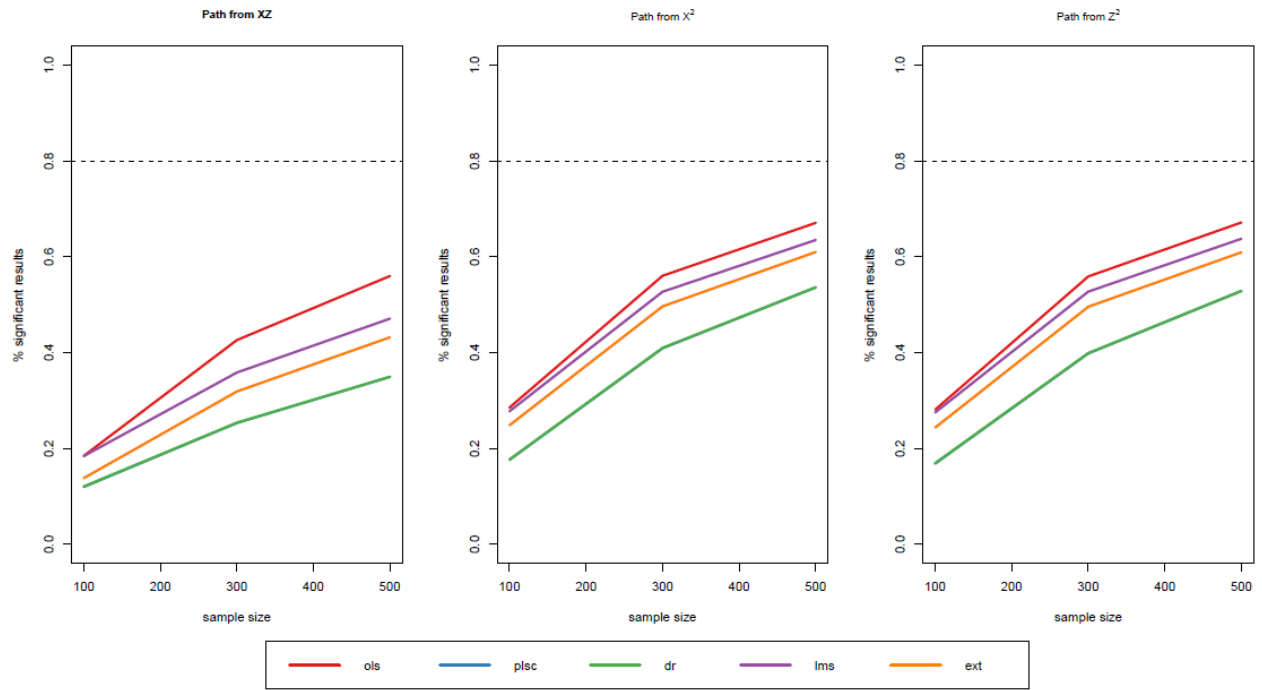


Figure 14. Statistical Power over Sample Size

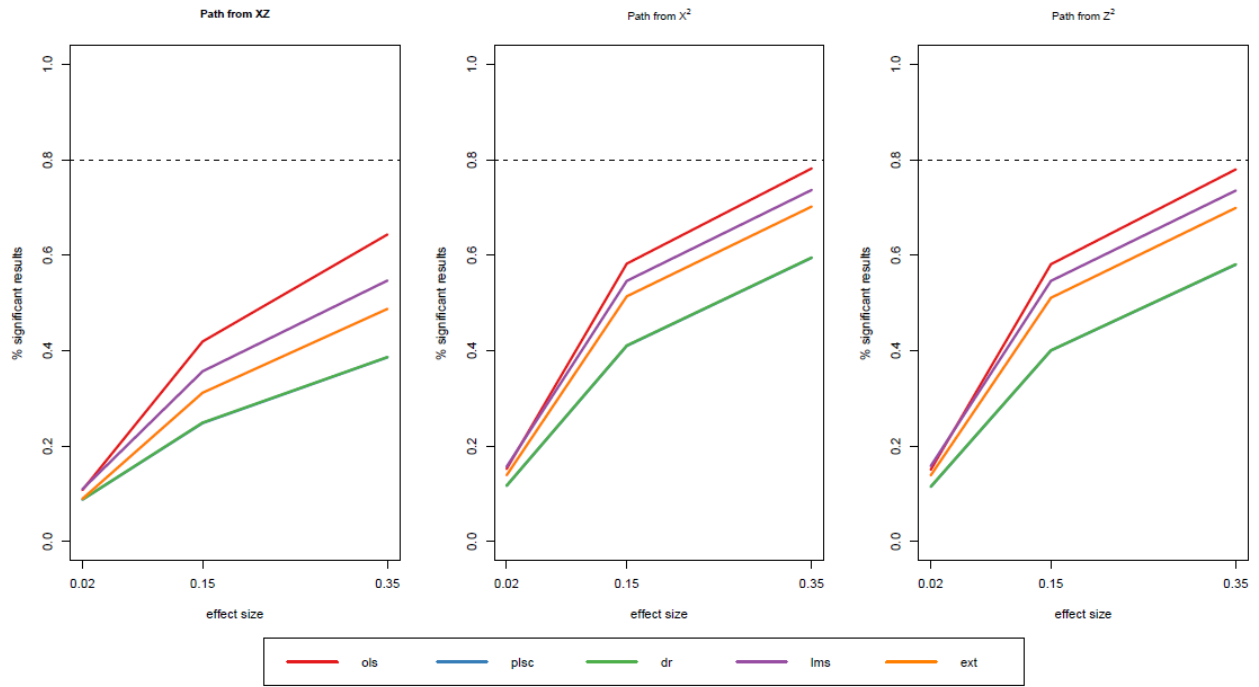


Figure 15. Statistical Power over Effect Size (f^2)

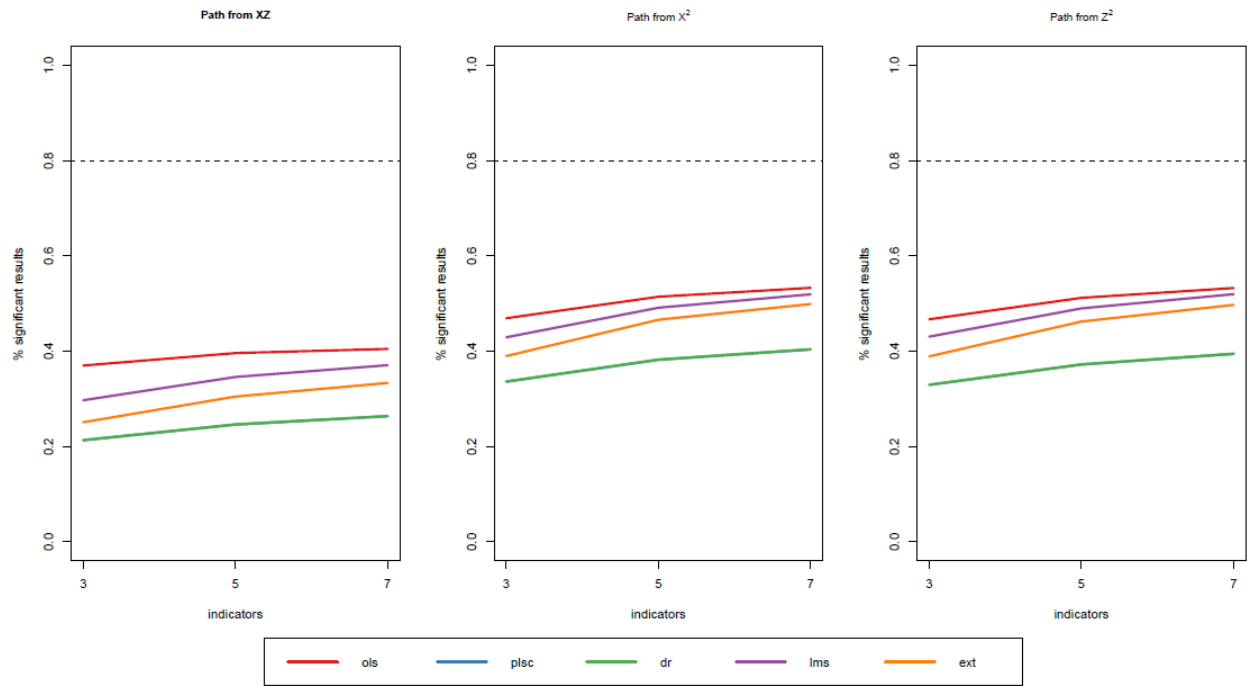


Figure 16. Statistical Power over Number of Indicators

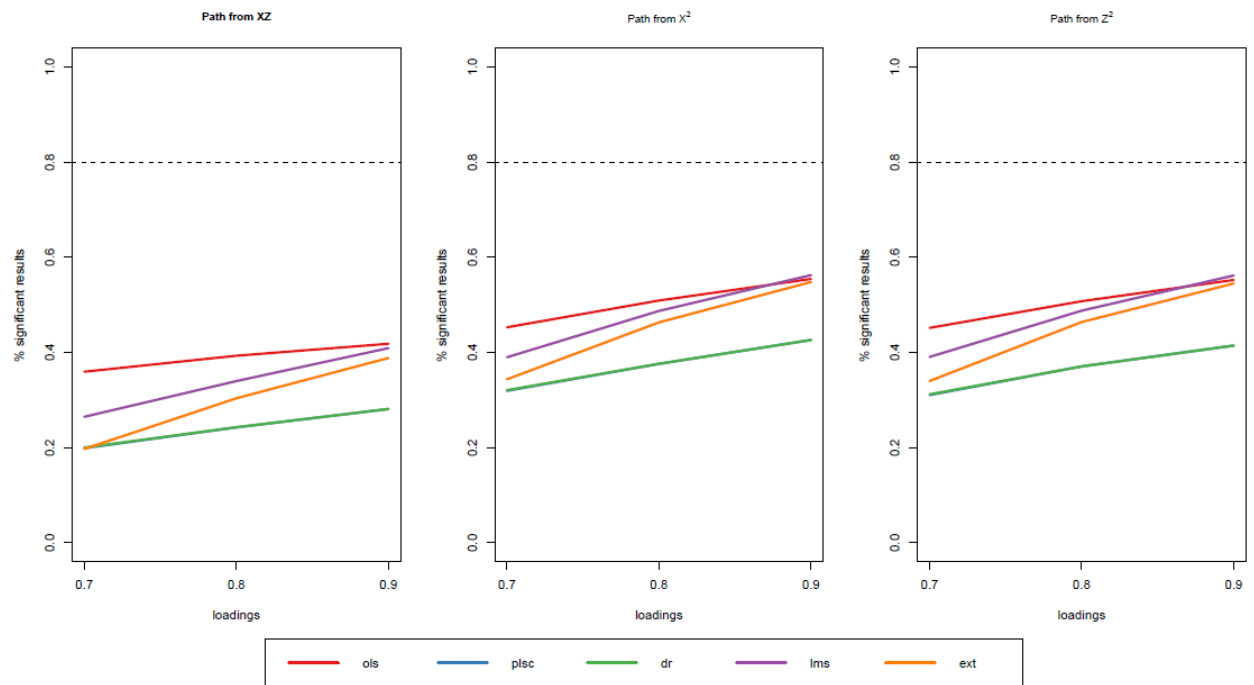


Figure 17. Statistical Power over Loading Strength

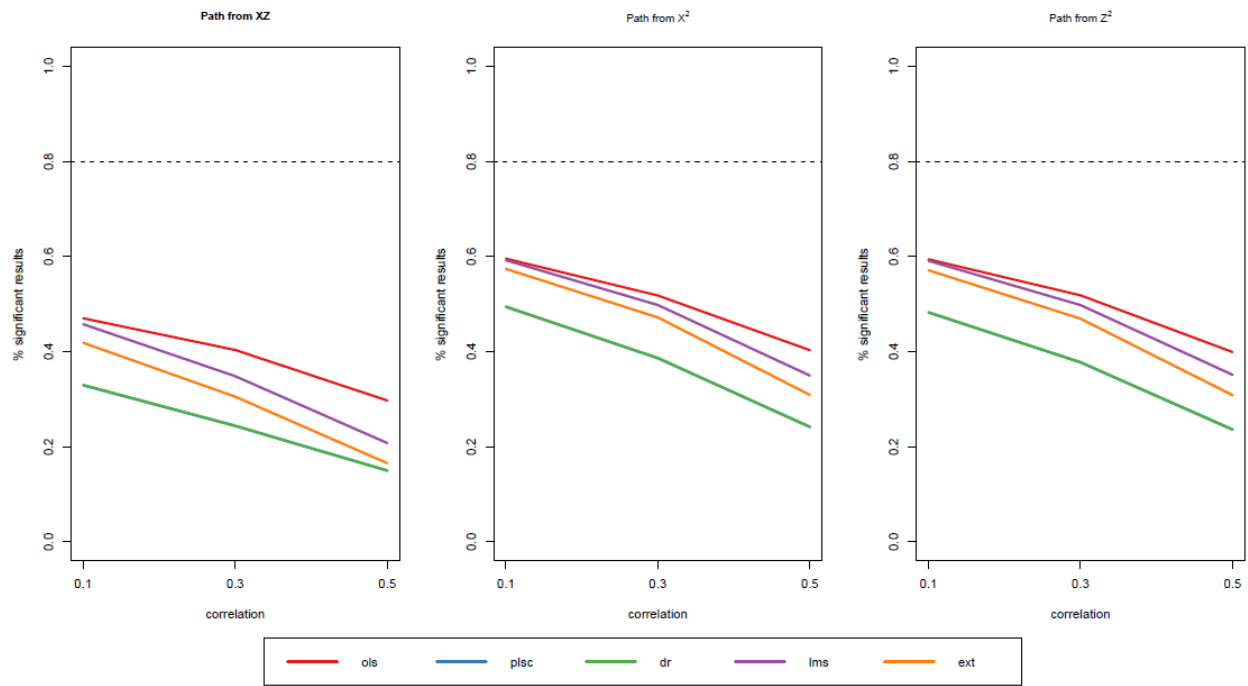


Figure 18. Statistical Power over Latent Variable Correlation

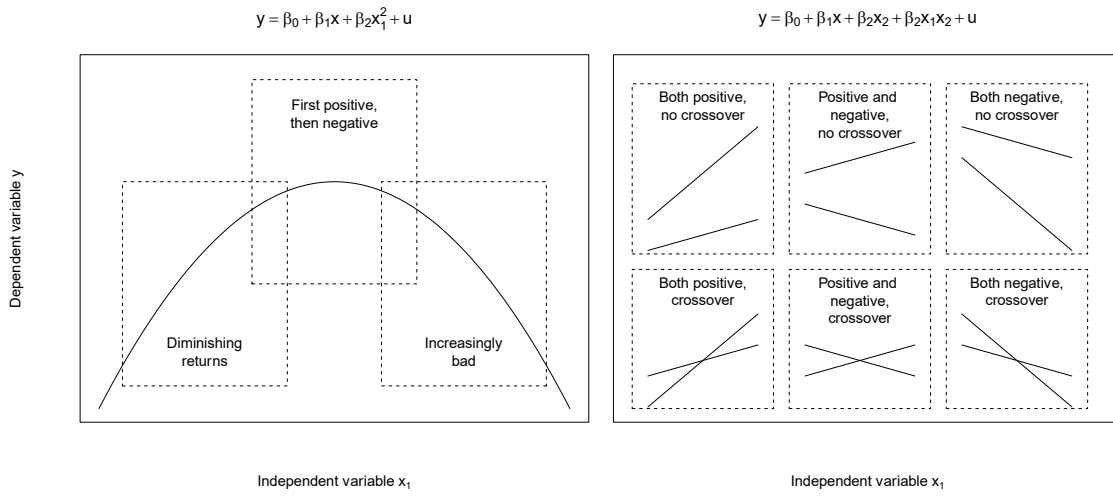
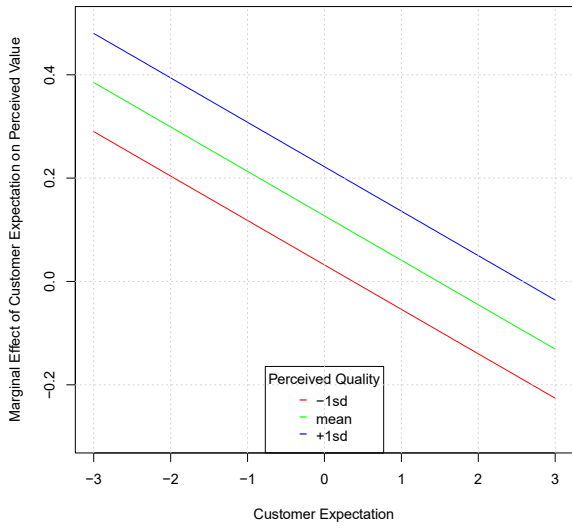


Figure 19 Interpretation of interaction and quadratic effects

OLS Marginal Effects



LMS Marginal Effects

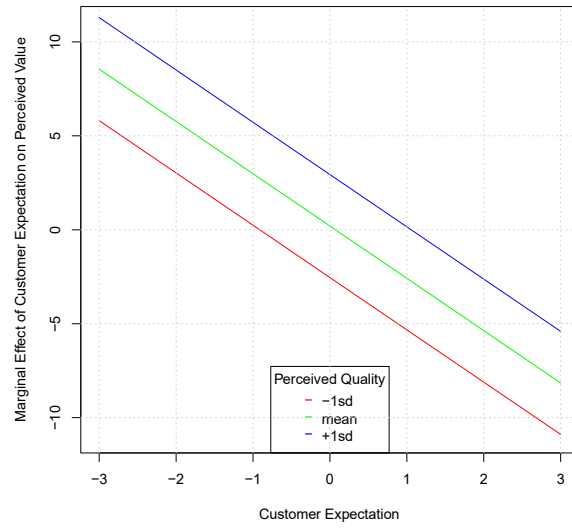


Figure 20. Marginal Effect Plots

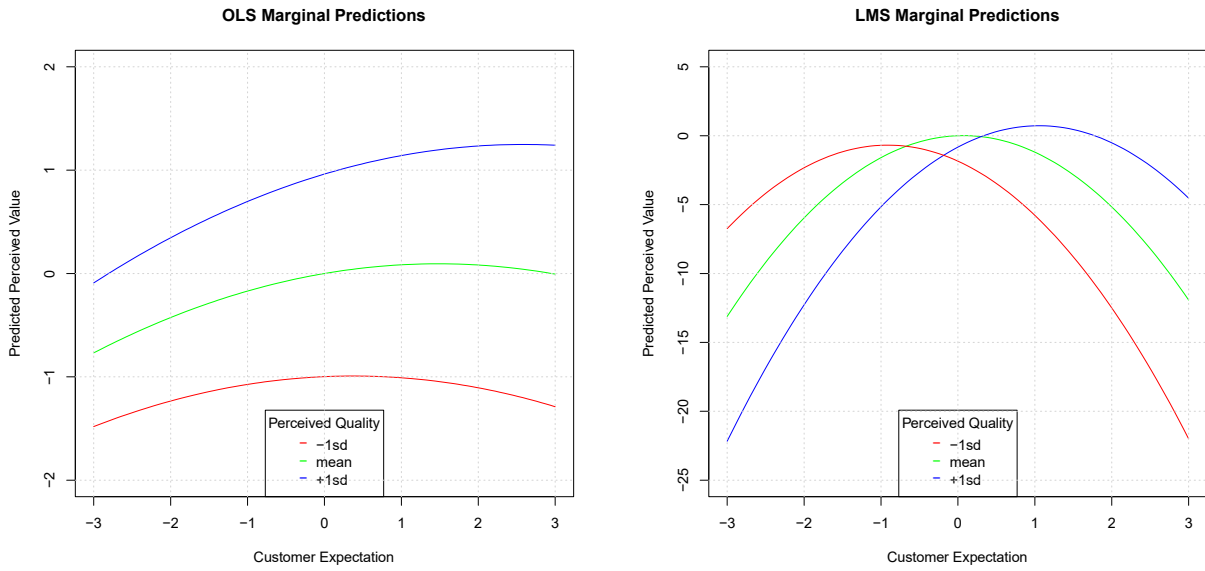
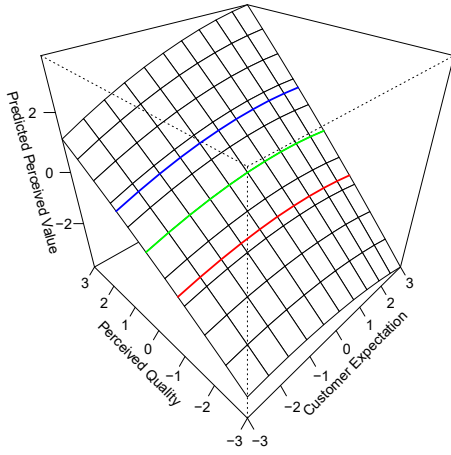


Figure 21. Marginal Prediction Plots

OLS Response Surface



LMS Response Surface

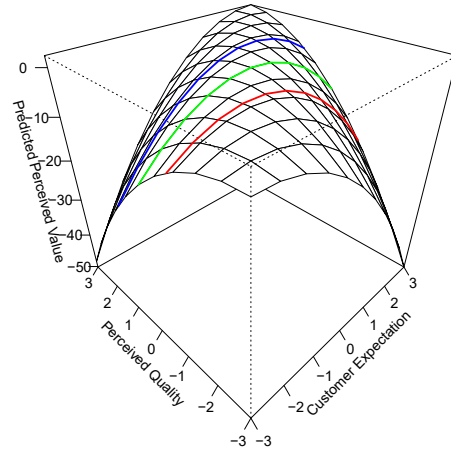


Figure 22. Response Surface Plots

Notes

¹ For example, in their study of the role of IT knowledge in venture capitalist overconfidence with regards to IT investments, Singh et al (2015) examined a research model which included interactions between quadratic terms and other predictors (e.g., [VC IT Knowledge]² × VC Successes) of the dependent variable.

² The full list of studies included in Table 1 is available from the first author upon request.

³ We do not endorse the use of reliability cutoffs and treating reliability as something that measures either have or do not have. While the cutoffs are commonly used in IS research, adjusting for measurement error, e.g. by using techniques presented in our study, is a much more appealing alternative because it allows researchers to assess what the estimated level of reliability means for the study results.

⁴ We note that methodologically both centering (Dalal & Zickar, 2012) and standardization (Bring, 1994; King, 1986) may not be as useful as commonly thought, and may in fact be problematic. We return to these issues in the discussion.

⁵ We note that Aguirre-Urreta and Rönkkö (2018) and many others (e.g., Cortina & Dunlap, 1997; Cortina & Landis, 2011; Cumming, 2014) point out that confidence intervals should not be used to make a yes/no decision by comparing the endpoints against a cutoff, but instead the endpoints should be interpreted. While we fully endorse this recommendation, in the context of a simulation study where the results must be comparable between techniques, this approach was not feasible.

⁶ Full results for all paths and conditions and for each statistical technique are available from the first author upon request.

⁷ Given the binary decision of the significant/not significant decision, it is possible to construct a confidence interval for $\alpha = 0.05$, given by $\alpha \pm z_{(1-\frac{\alpha}{2})} \sqrt{\alpha(1-\alpha)/k}$ where k is the number of replications (1,000 in this research). Under these conditions, a 95% confidence interval for the 5% alpha rejection rate is [0.036, 0.064].

⁸ We do not attach any particular theoretical significance to modeling the relationship as a second-order polynomial of the independent variables but do note that we find significance for the higher-order coefficients, indicating the possibility that a polynomial specification may indeed be appropriate. Instead, our goal here is simply to demonstrate the construction and interpretation of various plots from the results of a polynomial model.

⁹ A researcher employing OLS, would have concluded that there were no interaction or quadratic effects, whereas a researcher using LMS would have found two quadratic effects and an interaction. This difference further underscores the importance of employing the techniques recommended for the analysis of polynomial regression models when measurement error is present in the independent variables.