Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

Advances in Methods, Theories, and Philosophy

Dec 11th, 12:00 AM

Predictive Validation of Interaction Terms in PLS-SEM

Nicholas Danks
Trinity College Dublin, nicholas.danks@tcd.ie

Soumya Ray
National Tsing Hua University, soumya.ray@iss.nthu.edu.tw

Predictive Validation of Interaction Terms in PLS-SEM

Short Paper

Nicholas P. Danks
Trinity College
D02PN40 Dublin, Ireland
nicholas.danks@tcd.ie

Soumya Ray
National Tsing Hua University
Hsinchu 30013, Taiwan, R.O.C.
soumya.ray@iss.nthu.edu.tw

Abstract

The use of interaction terms in partial least squares structural equation modeling (PLS-SEM) risks overfitting models to small samples and producing poor out-of-sample generalizability. But the added complexity of interactions in PLS-SEM is not captured by in-sample fit metrics, and we propose that interaction terms in PLS-SEM should be assessed by out-of-sample methods and metrics. However, out-of-sample predictive methods like PLSpredict do not yet account for interaction terms. We start by providing a formal procedure for generating out-of-sample predictions from such models. We then empirically demonstrate that interactions produce far higher Type I error than that expected by researchers, and that out-of-sample predictive metrics indeed offer more accurate assessment of the validity of interaction terms for PLS-SEM. We also show that two-stage estimation of interactions is superior to other popular methods of operationalizing interactions in PLS-SEM, when the generalizability of interactions is of concern.

Keywords: Interactions, Overfit, Prediction, PLS-SEM

Introduction

Research models of latent–or emergent–constructs have become a mainstay of information systems research, and much of the management field at large. In particular, partial least squares structural equation modeling (PLS-SEM) has become increasingly popular as an approachable technique that fits models without the strong assumptions that parametric methods often require. These models have grown in complexity and maturity over the past decades, without a commensurate increase in the sample sizes being examined. To hone and help define the boundaries of these expansive theories, researchers frequently model contingencies in their models using interaction terms. Interactions help specify moderated relationships in which one construct's effects on an outcome is contingent on the level of another construct. It is operationalized by multiplying the two independent variables to create a product score.

However, higher-order terms like interactions add considerable complexity to SEM-type models because to introduce non-linearities they considerably add to the parametric complexity of the models, which in turn require greater statistical power to resolve (Goodhue et al. 2007). PLS-SEM models are already demonstrated to overfit to small samples and suffer poorer out-of-sample generalizability than simpler modeling schemes (Danks et al. 2023), and higher-order interaction terms are particularly mentioned as a worrying source of overfit. Yet there is little, if any, discussion or guidance on whether interaction terms should be included in research models estimated by PLS-SEM or other SEM methods. Given the above issues, we believe it is fair to ask how to validate that higher-order terms are adding generalizable value to SEM models, and when and how to properly include them.

Quantifying and resolving how the inclusion of higher-order terms in PLS-SEM models should be evaluated is not straightforward. Simple model fit metrics such as variance explained (R^2) favor overfit models (Danks et al. 2023) and so might be biased towards including spurious higher-order terms. Researchers often prescribe analyzing changes in model fit, sometimes operationalized as effect sizes, to determine the utility of higher-order terms (Henseler et al. 2012). But here again we are concerned about bias towards overfit, because model fit is of little value in the presence of overfit. Thus, interaction terms prove to be a blind spot for extant PLS-SEM fit analyses that solely emphasize model fit and statistical significance. We must look elsewhere to determine their utility and validity.

We propose that the questions about the inclusion and operationalization of higher-order terms in PLS-SEM are best addressed using out-of-sample predictive methods. Predictive methods can directly assess how models might perform on data they have never seen, and have been recently incorporated into PLS-SEM models using methods such as PLSpredict (Shmueli et al. 2016). But to do so, raises a few challenges. We must first produce a method to generate predictions from PLS-SEM models containing interactions because the PLSpredict procedure does not outline how to handle multiplicative terms and thus no software currently implements a formally scrutinized method for doing so. Second, we must validate that out-of-sample assessments of predictive power yield more useful information about the generalizability of interactions than in-sample fit based assessments like R^2. And third, we must revisit the question of how best to operationalize interaction terms in PLS-SEM—a few options abound and we suspect some might be more prone to overfit than others and so more likely to yield spuriously significant interactions. In doing so, we hope to provide a fresh perspective on the utility and validity of PLS-SEM models with interactions. We do so by raising awareness of the dangers of interaction terms in SEM modeling methods like PLS-SEM, and contribute an initial prediction-oriented method to assess the added value of interactions.

The Challenge of Interactions in SEM

The use of interactions to model moderation is not a simple affair in structural equation modeling, where interaction terms must be modeled as multi-item constructs in their own right. In traditional covariance-based SEM (CB-SEM) interaction terms are operationalized by taking products of all pairwise combinations of items of the independent and moderating constructs and using those products as items for the interaction term. PLS-SEM can also employ this *product indicator* approach, but more powerful approaches been developed (Henseler et al. 2012): *orthogonalized* product-terms approach, where item products are regressed over their original item pairs and the residual of the regression is used as the new product items; and the *two-stage* approach, where a model without any interaction is estimated so that the fitted construct scores of independent and moderator constructs can be retrieved and used to create a simpler, single product score.

The rising use of interactions in scientific research models is not unique to information systems but is rather a cross-disciplinary trend. And it has caught the wary eye of methodologists (Altman and Matthews, 1996). The theory-driven, empirical research tradition of SEM demands a hypothetico-deductive approach where the expectations for relationships are first justified before being confirmed against data using statistical tests. Post hoc justification of potential relationships discovered in an exploratory mode are strongly discouraged because spurious correlations can exist in any sample. The chosen significance level (e.g., α = 5%) of statistical tests dictates that a given percentage of relationships can be expected to be significant in a sample even if these relationships do not exist in the larger population.

Now let us consider an SEM model with ten exogenous constructs. In such a model, there might be a 5% expectation that any one of the paths might be spuriously significant even if it does not exist in the population. In this example, however, the 5% Type I error rate corresponds to less than a single expected spurious path, and so parsimonious models might be robust to minor opportunism by researchers who over-parameterize their models. But the situation becomes riskier when interactions are also considered: with ten exogenous constructs there are ${}_{10}C_2 = 45$ possible interactions, of which one or more spuriously significant findings are nearly guaranteed. Any opportunistic search for interactions in even such moderately sized models threaten the basis of empirical studies. The conventional wisdom on statistical testing already dictates that authors would be wise to avoid looking for interactions that were not strongly expected and justified in advance (Altman & Matthews, 1996). Reviewers and editors should already be wary of opportunism when presented with significant interactions in any regression type models.

But models with latent or composite constructs require many more parameters than ordinary regression models to represent measurement theory. For example, constructs might each be measured by three manifest items collected from either survey questions, observable metrics, or public information such as market indices. In our earlier example of a model with ten exogenous constructs and an endogenous construct, our measurement model requires 33 additional item weight or loading parameters to relate measurement items to constructs, on top of the ten path coefficients needed by the structural model to relate independent variables to the dependent variable. So, where an ordinary regression model might have only ten coefficients to estimate a model of the same structural complexity, the corresponding PLS-SEM model using multi-item constructs would need at least 43 estimands. CB-SEM models need to estimate even more parameters to represent the inter-construct and inter-item associations that are constrained in PLS-SEM. The added parameters that SEM models must estimate gives them more opportunities than regression models to overfit to the data and produce spurious findings. And our concern lies with the inclusion of interaction terms that require a multiplicative number of items to represent moderation between multi-item constructs.

We sought to demonstrate the added dangers of overfit to detecting interactions in SEM by means of a comparative simulation exercise. We constructed the SEM model described earlier with 10 exogenous constructs that relate to a single endogenous construct. We used the survey data and the associated model described and made publicly available by Danks et al. (2023). To this model we iteratively added all 45 possible interactions between the exogenous constructs. Surprisingly, 5 of the 45 interactions (11%) were found to be significant at $\alpha = 5\%$. While one or more of the significant interactions might be genuine, and empirical researchers might be tempted to claim multiple of them, we recall the warnings of methodologists who caution against any significant interactions that are not drawn from robust theorization and prior expectations (Altman & Matthews, 1996).

We contrasted these results of the PLS-SEM model to an ordinary multiple regression formulation of the same structural complexity with ten path coefficients, but where all measurement parameters were shed by simply averaging the respective measurement items of each construct. We added to the regression model the 45 possible interactions between the ten independent variables:

$$BI = \beta_0 + \beta_1 PE + \beta_2 EE + [\ldots] + \beta_{11} PE{*}EE + [\ldots] + \varepsilon. \tag{1}$$

On the contrary, when estimating the multiple regression model we find only 2 of the 45 interactions (4.44%) to be significant at $\alpha = 5\%$. The stark contrast between the results of the PLS-SEM model and its multiple regression counterpart provides preliminary evidence of how the extra specifications of an SEM model overfits to interactions, and yields excess Type I errors. This demonstration puts significance- and fit-based methods to detect interactions in further doubt. We even caution against construct-specific fit measures such as R^2 and its simple derivatives such as adjusted-R^2, which entail analyzing in-sample fit and are thus vulnerable to overfit.

Predictions of PLS-SEM Models Involving Interactions

A key challenge that is encountered when applying prediction to SEM is that actual (or at least estimated) factor scores are needed to compare against predictions (Rigdon 2012; Danks et al. 2023). Because the scores of latent factors in CB-SEM models are indeterminate, there is no extant approach for generating construct-level predictions from CB-SEM models. We thus focus on PLS-SEM, in which construct scores are determinable and can be derived from estimated item weights. Methodologists have already demonstrated how to generate predictions from PLS-SEM models using a procedure called PLSpredict (Shmueli et al. 2016). However, generating predictions from models containing interaction terms is not straightforward, and PLSpredict did not consider how to handle interaction terms. In this study, we contribute methods by which to generate predictions from PLS-SEM models containing interactions.

Procedure for generating point predictions from PLS models with interactions

Each modeling approach for interaction terms in PLS-SEM requires its own custom procedure for generating predictions. Procedures for generating predictions for product-indicator and orthogonalized approaches are available online along with examples at https://github.com/sem-in-r/moderator_predict. The procedure to generate predictions involving interactions modeled and estimated by the two-stage approach is shown below. We focus on the two-stage approach because, as we will see, it generates the most generalizable predictions and should be the first choice of researchers seeking predictive validation of interactions. The procedure below estimates training weights from a training sample and then generates predictions on a new sample, which can either be a holdout set for validation purposes or data from a new context or time for practical purposes.

1. Identify training (x_{is}) and holdout (x_{oos}) sets of cases. Holdout data can be new data or data that has been partitioned during cross validation.
2. Estimate parameters of the **first** and **second** stage models using training data only.
   a. Retain initial descriptive statistics of the training data for the **first stage** (mean $\widehat{m1}_{is}$ and standard deviation $\widehat{s1}_{is}$), and for the **second stage** (mean $\widehat{m2}_{is}$ and standard deviation $\widehat{s2}_{is}$)
   b. Retain estimated parameters for the **first stage** measurement weights ($\widehat{w1}_{is}$), loadings ($\widehat{l1}_{is}$), and structural path coefficients ($\widehat{B1}_{is}$), and for the **second stage** measurement weights ($\widehat{w2}_{is}$), loadings ($\widehat{l2}_{is}$), and structural path coefficients ($\widehat{B2}_{is}$).
3. Generate the holdout indicator data for the non-linear term
   a. Standardize holdout data from step 1 using first stage model standard deviation $\widehat{s1}_{is}$ and mean $\widehat{m1}_{is}$.
   b. Predict the construct scores of exogenous constructs using holdout data (x_{oos}) from step 3.a and the first stage training measurement weights ($\widehat{w1}_{is}$):
   $$\hat{X} = x_{oos}.\widehat{w1}_{is}.$$
   c. Generate the holdout non-linear term indicator score by multiplying the construct scores of the exogenous and moderator variables from 3.b.
   d. Append the holdout data (x_{oos}) with the non-linear term indicator score calculated in 3.c.
4. Standardize holdout data from step 3 using second stage model standard deviation $\widehat{s2}_{is}$ and mean $\widehat{m2}_{is}$.
5. Predict exogenous construct scores from outer weights:
   Predict the construct scores of exogenous constructs using holdout data from step 4 and the second stage model measurement weights ($\widehat{w2}_{is}$):
   $$\hat{X} = x_{oos}.\widehat{w2}_{is}.$$
6. Predict the endogenous construct scores:
   Multiply the predicted construct scores ($\hat{X}$) by second stage structural paths ($\widehat{B2}_{is}$):
   $$\hat{Y} = \hat{X}.\widehat{B2}.$$
7. Predict the indicator scores of endogenous constructs:
   Multiply the predicted construct scores ($\hat{Y}$) with the second stage measurement loadings ($\widehat{l2}_{is}$):
   $$\hat{y} = \hat{Y}.\widehat{l2}_{is}.$$
8. Unstandardize predictions.
   Use the second stage standard deviation $\widehat{s2}_{is}$ and mean $\widehat{m2}_{is}$ to bring the predictions back to the original scale. Multiply each predicted observation by its corresponding standard deviation and add its corresponding mean.

Guidelines for creating the interaction term for non-linear effects in PLS-SEM are well described in the methodological literature (Becker et al. 2018; 2023, Henseler and Chin, 2010, and Henseler, Fassott, Djikstra, and Wilson, 2012). Henseler and Chin (2010) as well as Henseler et al. (2012) pointed out that the orthogonalizing and product indicator approaches explain a significantly and substantially higher proportion of variance in the outcome construct as compared to the two-stage approach. In addition, when the primary concern is minimizing estimation bias, the orthogonalizing approach should be preferred as it performs best in terms of point accuracy. When the goal is to achieve statistical power or to identify the statistical significance, the two-stage approach is preferred as opposed to the other two approaches. Importantly, their simulations show similar results for moderation and quadratic effects.

Becker et al. (2018) later presented a more extensive simulation which investigated the relative efficacy of the three approaches in terms of parameter recovery. Their simulation study demonstrated that the two-stage approach clearly outperforms the other approaches in terms of parameter recovery, and operationalizes the interaction term in a simpler way with substantially less estimated parameters. Importantly, the two-stage approach is the most versatile approach as it can be used regardless of whether the exogenous construct is measured reflectively or formatively (Becker et al., 2023).

However, it is important to note that these simulations emphasize only the in-sample prediction accuracy (explanatory power in terms of R^2 without considering out-of-sample prediction accuracy. Although we recognise that the primary concern of researchers is the significance and magnitude of parameters that in turn allow for inferential conclusions on hypothesized relationships, we argue that the decision to include non-linear effects (or indeed which estimation approach to employ) should in part be motivated by the efficacy in terms of out-of-sample generalizability and a balance of fit versus prediction (overfit).

We expect that the complexity of both the product-indicator and orthogonal approaches will likely result in overfitting, yielding poorer out-of-sample prediction accuracy as compared to the simpler two-stage approach. James et al. (2013 p. 144) note that: "The higher the ratio of parameters p to number of samples n, the more we expect overfitting to play a role. Both the product-indicator and orthogonalizing approaches require the multiplication of product indicators for each of both the exogenous and moderator constructs. Consider again our example mentioned above, where the direct-effects model has 43 estimands. For each additional non-linear term added to the model a minimum of 16 additional estimands are introduced. We thus propose that two-stage might yield the most robust results.

Simulation Study

This simulation study explores the utility of prediction metrics in generating additional evidence to support the validation of interaction terms in SEM. We focus our efforts on PLS-SEM as, to the best of our knowledge, there is no extant solution for generating predictions for CB-SEM models. We further compare the predictive performance of the three approaches to estimating the interaction term in order to understand their relative advantages and shortcomings. The two-stage approach has been shown to perform best at parameter recovery and demonstrates higher power in comparison to the orthogonalizing and product-indicator approaches (Becker et al., 2018).

We apply the method outlined by Becker et al. (2018) in our data simulation process. In this data generation model (DGM), we first generate the construct scores to include the non-linear relation and then generate the indicator scores from the construct scores. The data generation model consists of one antecedent construct (X1), one moderator construct (X2), one endogenous construct (Y), and an interaction term construct (X1*X2; Fig 1 panel a) that produces the non-linear moderating effect.

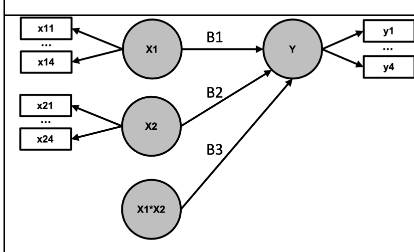
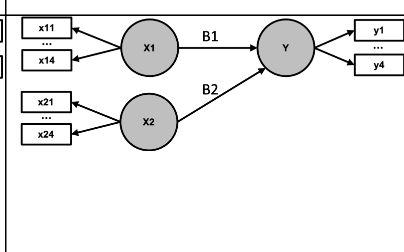
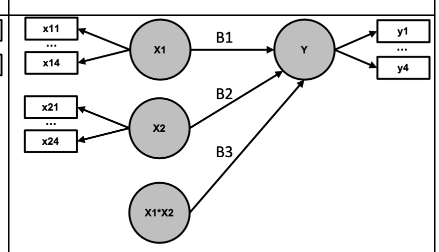
1.a) Data generation model	1.b) Main effects model	1.c) Moderation model
		
We vary the B3 interaction coefficient across 4 levels: 0 (no interaction), 0.05, 0.1, and 0.2.	When DGM interaction B3 = 0, this model is correctly specified, otherwise it is misspecified	When DGM interaction term B3 > 0, this model is correctly specified, otherwise it is misspecified

Figure 1: Simulation models

We vary B3 in the DGM to be both zero and non-zero in order to take into account the condition where the interaction term is specified to have no effect. When B3=0, we consider the DGM to be a non-moderation condition and specify it as such in our tables and results. Having this condition allows us to compare the performance of the methods when no moderation is present in the DGM.

For the data generation we use the R framework and the MASS library (R Core Team, 2014; Venables & Ripley, 2002). We apply the data generation process to generate a single population dataset with 100,000 observations. In each iteration of the simulation, we then sample from that population a subsample with n observations. For each subsample, we conduct leave-one-out cross validation (LOO-CV) and apply each of the three proposed algorithms for generating predictions from moderated PLS-SEM models outlined above, and a naive model with no interaction term. Then using both the in-sample ($\hat{Y}_{in}$) and out-of-sample predictions ($\hat{Y}_{out}$) we calculate and report the in-sample mean square error (MSE_{in}), and out-of-sample MSE (MSE_{out}) for the focal outcome construct (Danks et al., 2023). We estimate the PLS models using the SEMinR package (Ray et al., 2022) for the R Statistical environment.

The goals of this simulation study are fourfold: 1) We first aim to quantify the predictive performance of the three algorithms, and generate guidelines for their application. 2.) We wish to identify whether predictive metrics can reliably identify when the DGM does not include an interaction effect, akin to a true negative in traditional statistical testing; 3) We wish to evaluate whether prediction metrics can reliably identify when the DGM includes an interaction effect, akin to a true positive in a traditional statistical testing. 4) Evaluate whether prediction power metrics can provide reliable evidence to the nature of the DGM.

Following similar methodological work by Sharma et al. (2019, 2021) and guided by the meta-analysis of PLS-SEM in information systems research conducted by Hair et al. (2017), we manipulated the following experimental conditions, which correspond to the conditions commonly encountered in applied research (Nitzl, 2016; Ringle et al., 2020):

- Replications: 50
- Five conditions of sample size: 150, 250, 500, 1000, and 1500
- Four conditions of interaction term effect size B3: 0, 0.05, 0.1, 0.2,
- Three conditions of direct effects B1 and B2: 0.1 and 0.1, 0.25 and 0.25, 0.4 and 0.25
- Three indicator loading patterns with different levels of average variance extracted (AVE): High AVE loadings: (0.9, 0.9, 0.9, and 0.9), Moderate AVE loadings: (0.8, 0.8, 0.8, and 0.8), and Low AVE loadings: (0.7, 0.7, 0.7, and 0.7).

DGM	Interaction Effect (β3)	Metric	Two-stage Interaction Model	PI Interaction Model	Orthogonal Interaction Model	Main Effects Model
Non-moderation (β3 = 0)	0.00	MSE_{in}	.6997	.6953	**.6940**	.7013
		MSE_{out}	.7150	.7221	.7261	**.7126**
		Overfit%	2.2 %	3.9 %	4.6 %	**1.6 %**
Moderation (β3 > 0)	0.05	MSE_{in}	.6622	.6584	**.6572**	.6683
		MSE_{out}	**.6782**	.6828	.6847	.6806
		Overfit%	2.4 %	3.7 %	4.2 %	**1.8 %**
	0.10	MSE_{in}	.6191	.6157	**.6153**	.6417
		MSE_{out}	**.6337**	.6356	.6360	.6525
		Overfit%	2.4 %	3.2 %	3.4 %	**1.7 %**
	0.20	MSE_{in}	.5151	**.5114**	.5121	.6114
		MSE_{out}	**.5294**	.5300	.5308	.6236
		Overfit%	2.8 %	3.6 %	3.7 %	**2.0 %**
Moderation Mixed	Grand Total	MSE_{in}	.5988	.5952	**.5949**	.6405
		MSE_{out}	**.6137**	.6161	.6172	.6522
		Overfit%	2.5%	3.5 %	3.8 %	**1.8 %**

Table 1. Simulation results by interaction effect size

Notes: MSE averaged across all simulation conditions; ME Model: Main effects model, in: in-sample, out: out-of-sample, MSE: mean square error, Overfit%: overfit ratio. Bolded values indicate lowest error (highest prediction power).

Consulting Table 1, a striking result is that the moderated models yield an improved model fit (MSE_{in}) over the ME model under all simulation conditions–even when the ME model is correctly specified ($\beta 3 = 0$). When consulting only the in-sample metric a researcher might not identify potential misspecification. In contrast, the out-of-sample metric (MSE_{out}) correctly identifies the ME model with the lowest MSE_{out} in the non-moderation condition. This highlights the dangers of considering only in-sample metrics when evaluating a model's fit to the data. When we zoom in on the moderation approaches we find that the orthogonal and product-indicator approaches consistently have the lowest is-mse (best model fit), but the worst predictive power. These approaches have a consistently higher overfit ratio–a phenomenon which is likely due to the high parameterization of these approaches. The two-stage approach yields the best balance of fit and prediction while achieving the best predictive power. Furthermore, it confirms that out-of-sample prediction is an effective way to identify moderation.

Next, we consider the accuracy of the two-stage approach and the MSE metric for correctly identifying the true DGM – that is for identifying when the "true" model does and does not include an interaction. We consider this comparison in the same light as a hypothesis test – where the null hypothesis is that no moderation is present (H_{NULL}: $\beta 3 = 0$) and the alternative is that moderation is present (H_{ALT}: $\beta 3 \neq 0$). We then analyze the results from the above simulation, identifying for each DGM condition (DGM includes a moderator or not) whether the MSE_{out} preferred the main effects or moderated model with lower predictive error. We then calculate percentages for each and report below in the confusion matrix (Table 2).

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Data generation model (Null hypothesis) is</td></tr>
<tr><td>TRUE</td><td>FALSE</td></tr>
<tr><td rowspan="2">Decision about Null Hypothesis</td><td>Don't Reject</td><td>Correct inference / true negative
(1-α)
MSE_{in}: 25.3 %
MSE_{out}: 84.6%</td><td>Type II Error / false negative
(β)
MSE_{in}: 3.9 %
MSE_{out}: 15.6 %</td></tr>
<tr><td>Reject</td><td>Type I Error / false positive
(α)
MSE_{in}: 74.7%
MSE_{out}: 15.4%</td><td>Correct Inference / true positive
(1-β)
MSE_{in}: 96.1 %
MSE_{out}: 84.4 %</td></tr>
<tr><td colspan="4">Table 2: Alpha and Power for two-stage out-of-sample MSE
Notes: MSE: mean square error; in: in-sample; out: out-of-sample</td></tr>
</table>

Overall, we find that using MSE_{out} to identify interactions across all simulation conditions yields a true negative rate of 84.6% ($\alpha = 0.154$) and a true positive rate of 84.4% ($\beta = 0.156$) indicating that MSE_{out} is an effective metric to identify interaction terms. When we consider a model and dataset with qualities more typical of the IS literature–with moderate AVE (0.8), sample size (500), and effect size (0.2)–the method is even more accurate. The true negative rate jumps to 92.5 % ($\alpha = 0.075$) and the true positive rate rises to nearly 98% ($\beta = 0.02$). Thus, when a model meets good quality considerations, the method is very accurate at correctly identifying the interaction.

Future Work

We have thus far uncovered preliminary evidence that SEM estimation methods such as PLS-SEM are far more prone to finding spuriously significant interactions than ordinary linear regression. We believe that fit-based metrics of model utility. However, we must carry out broad simulations of varying conditions and metrics to demonstrate and confirm these suspicions. We have also started to resolve our question of how to validate interactions in PLS-SEM by examining how out-of-sample predictive metrics respond to nonlinearities. Our initial findings are that predictive metrics are more conservative than in-sample metrics, in that they are less likely to find that spurious interactions have generalizable value. However, here too we hope that future work expands on our initial findings in more interesting ways. There are a plethora of predictive metrics and we must find a defensible subset to compare. Moreover, we hope to arrive at a validation framework that also allows researchers to safely examine non-hypothesized interactions with the hope of discovering fruitful contingencies for future research. In doing so, we hope to live up to the emerging vision of the information systems field, where computational methods can inform future discoveries and theories. Finally, interactions can be part of larger theoretical patterns, such as moderated mediation, that need to be re-examined in light of this study.

References

Altman, D. G., & Matthews, J. N. (1996). Statistics Notes: Interaction 1: heterogeneity of effects. *BMJ*, *313*(7055), 486.

Becker, J. M., Cheah, J. H., Gholamzade, R., Ringle, C. M., & Sarstedt, M. (2023). PLS-SEM's most wanted guidance. International Journal of Contemporary Hospitality Management, 35(1), 321-346.

Becker, J. M., Ringle, C. M., & Sarstedt, M. (2018). Estimating moderating effects in PLS-SEM and PLSc-SEM: Interaction term generation* data treatment. *Journal of Applied Structural Equation Modeling*, 2(2), 1-21.

Danks, N. P., Ray, S., & Shmueli, G. (2023). The Composite Overfit Analysis Framework: Assessing the Out-of-Sample Generalizability of Construct-Based Models Using Predictive Deviance, Deviance Trees, and Unstable Paths. *Management Science,* forthcoming – available online Ahead of Print at: https://pubsonline.informs.org/doi/full/10.1287/mnsc.2023.4705

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

Hair, J., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial management & data systems*.

Henseler, J., Fassott, G., Dijkstra, T. K., & Wilson, B. (2012). Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. *European Journal of Information Systems*, *21*, 99-112.

Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. Structural equation modeling, 17(1), 82-109.

Henseler, J., Fassott, G., Dijkstra, T. K., & Wilson, B. (2012). Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. European Journal of Information Systems, 21, 99-112.

James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning (Springer, New York).

Nitzl, C. (2016). The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development. *Journal of Accounting Literature*.

R Core Team. (2023). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ray S, Danks NP, Calero Valdez A, (2022) seminr: Building and Estimating Structural Equation Models. R package version 2.3.2. Accessed February 7, 2023, https://cran.r-project.org/web/ packages/seminr/.

Rigdon EE (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning* 45(5–6): 341–358.

Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, 31(12), 1617-1643.

Sharma P, Sarstedt M, Shmueli G, Kim KH, Thiele KO (2019) PLS-based model selection: The role of alternative explanations in information systems research. *J. Assoc. Inform. Systems* 20(4):4.

Sharma PN, Shmueli G, Sarstedt M, Danks N, Ray S (2021) Prediction-oriented model selection in partial least squares path modeling. *Decision Science* 52(3):567–607.

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. 2016. The elephant in the room: Predictive performance of PLS models. *Journal of business Research*, *69*(10), 4552-4564.

Venables , W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S (4th ed.). New York: Springer.