PLS-Based Model Selection: The Role of Alternative Explanations in Information Systems Research

Article *in* Journal of the Association for Information Systems · April 2018
DOI: 10.17705/1jais.00538

Journal of the Association for Information Systems (2019) **20**(4), 346-397
doi: 10.17005/1.jais.00538

RESEARCH PAPER

ISSN 1536-9323

PLS-Based Model Selection: The Role of Alternative Explanations in Information Systems Research

Pratyush Nidhi Sharma[1], Marko Sarstedt[2], Galit Shmueli[3], Kevin H. Kim[4] Kai Oliver Thiele[5]

[1]University of Delaware, USA, pnsharma@udel.edu
[2]Otto-von-Guericke University, Germany; Monash University, Malaysia, marko.sarstedt@ovgu.de
[3]National Tsing Hua University, Taiwan, galit.shmueli@iss.nthu.edu.tw
[4]University of Pittsburgh, USA
[5]Hamburg University of Technology, Germany, kai.oliver.thiele@arcor.de

Abstract

Exploring theoretically plausible alternative models for explaining the phenomenon under study is a crucial step in advancing scientific knowledge. This paper advocates model selection in information systems (IS) studies that use partial least squares path modeling (PLS) and suggests the use of model selection criteria derived from information theory for this purpose. These criteria allow researchers to compare alternative models and select a parsimonious yet well-fitting model. However, as our review of prior IS research practice shows, their use—while common in the econometrics field and in factor-based SEM—has not found its way into studies using PLS. Using a Monte Carlo study, we compare the performance of several model selection criteria in selecting the best model from a set of competing models under different model set-ups and various conditions of sample size, effect size, and loading patterns. Our results suggest that appropriate model selection cannot be achieved by relying on the PLS criteria (i.e., R^2, Adjusted R^2, GoF, and Q^2), as is the current practice in academic research. Instead, model selection criteria—in particular, the Bayesian information criterion (BIC) and the Geweke-Meese criterion (GM)—should be used due to their high model selection accuracy and ease of use. To support researchers in the adoption of these criteria, we introduce a five-step procedure that delineates the roles of model selection and statistical inference and discuss misconceptions that may arise in their use.

Keywords: Information Criteria, Partial Least Squares (PLS), Structural Equation Modeling (SEM), Model Selection, Model Selection Criteria, Monte Carlo Study

Stacie Petter was the accepting senior editor. This research article was submitted on October 3, 2014, and went through three revisions.

1 Introduction

Sociotechnical systems within the purview of information systems (IS) research are inherently complex due to intricate underlying causal interactions and processes. Models are formal quantitative representations of theories and hypotheses that are devised to offer *partial* explanations of such complex systems (Lauenroth, 2003). Researchers build models that serve distinct goals in varied settings of interest to *approximate* these processes by abstracting away the details. This may result in the existence of multiple models reflecting varied theoretical lenses, levels of development, assumptions, interpretations, contexts, or even current fads. While this diversity enriches the literature, it also creates the threat of fragmentation of the field due to the individual researchers' "parental affections" for particular models (Chamberlin, 1890; Grover, 2013). The role of the scientific process, and model comparison in particular, is to overcome the human biases and sharpen our understanding over time by identifying and selecting the best approximating model(s) explaining a phenomenon, thereby bringing the field together (Purcell, 1992).

Philosophers of science have long realized the importance of considering alternative explanations (i.e., models) when researching certain phenomena. For example, Popper (1959) argued that considering alternative explanations (or "possible causes") is a crucial step *prior* to any attempt at the "falsification" of a theory. While pondering over the question of why some fields saw faster scientific advances than others, Platt (1964, p. 350) reasoned that,

> *"The conflict and exclusion of alternatives that is necessary to sharp inductive inference has been all too often a conflict between men, each with his single Ruling Theory. But whenever each man begins to have multiple working hypotheses, it becomes purely a conflict between ideas".*

More recently, Nuzzo (2015) warned against cognitive fallacies that may lead researchers to make serious scientific errors, such as collecting evidence to support a specific hypothesis, not looking for evidence against it, and ignoring alternative explanations. To counter this, she calls researchers to explicitly consider *plausible* (i.e., motivated by theory) alternative explanations that are not just strawmen, but span models that offer theoretically justified alternatives for explaining the phenomenon under study.[1]

Alternative explanations can come in different forms and give rise to several models with different (or additional) antecedents and/or model relationships, all of which are plausible within the realm of the theoretical framework(s) under consideration. For example, researchers may derive alternative models from a single theory or multiple theories, such as Venkatesh, Morris, Davis, and Davis (2003) who relied on model comparison to benchmark the unified theory of acceptance and use of technology (UTAUT) model against alternatives. Similarly, Plouffe, Hulland, and Vandenbosch (2001) compared alternative models derived from the technology acceptance and the perceived characteristics of innovating frameworks to benchmark their explanatory power.

Alternative models may also emerge when considering theories in new contexts with unique variables and effects. In this vein, Johns (2006) and Alvesson and Kärreman (2007) note that new contexts can result in important changes in theories, such as rendering originally theorized relationships redundant or altering their magnitude, and/or creating new relationships by introducing new antecedents. For example, Venkatesh Thong, and Xu (2012) tailored the UTAUT to a consumer context by identifying additional constructs and relationships and compared their UTAUT2 model with the original model.

Finally, alternative models may be created when researchers seek to build conceptual bridges across related streams of inquiry to provide a holistic understanding of the phenomenon. For example, Wixom and Todd (2005) note that the technology acceptance and user satisfaction literatures evolved in parallel and engage in model comparisons to integrate them.

Thus, considering alternative models can facilitate the development of stronger theory by extending or challenging the assumptions of existing theory, integrating parallel streams of inquiry, or benchmarking against more accurate and generalizable alternatives (Grover, Lyytinen, Srinivasan, & Tan, 2008). In line with these notions, IS theorists have long called for comparing alternative models, such as Roberts and Grover (2009, p. 89) who note that "specifying alternative models, is useful in theory building because it gives the researcher alternative perspectives concerning the focal phenomena…. We recommend that IS researchers compare alternate a priori models to discover the model that the observed data support best".

The need for considering alternative models has also been stressed in partial least squares path modeling (PLS) (Wold, 1980), a widely used regression-based technique in IS and other fields that estimates relationships in path models with latent and manifest variables (Hair, Hollingsworth, Randolph, & Chong, 2017; Hair, Sarstedt, Ringle, & Mena, 2012; Khan et al. 2019; Lee, Petter, Fayard, & Robinson, 2011; Marcoulides & Saunders, 2006; Ringle, Sarstedt & Straub, 2012). For example, Rigdon, Sarstedt, and Ringle (2017, p. 13) recently stressed that "researchers should more routinely explore theoretically justified alternative models for explaining the phenomenon under study". Similarly, Gefen, Rigdon, and Straub (2011) advise that PLS researchers should more frequently engage in model comparisons. These recommendations follow the principles of factor-based structural equation modeling where model comparisons have been routinely performed since the 1980s (e.g., Anderson & Gerbing, 1988; Anderson, Gerbing, & Hunter, 1987; Gerbing & Anderson, 1988; Lin, Huang, & Wenig, 2017). In the same vein, Hermann Wold (1980, p. 70), the originator of PLS,

[1] We emphasize *plausible* because the total number of alternative explanations, plausible plus implausible, can be exceedingly large (Dill, 2013). The role of theory is paramount in judging whether such alternative explanations are plausible.

noted that model construction in PLS is an “evolutionary process”, which involves comparing alternative models, each grounded in theory (e.g., Wold, 1982, 1985). However, none of these authors offer any perspective on exactly *how* to empirically compare alternative models and select a model that is best supported by the data within the PLS framework.

As we will show later, despite the central role played by alternative explanations in the scientific method, model selection is rarely performed in PLS-based studies. We will also show that PLS criteria (i.e., criteria used to assess the quality of the structural model—e.g., R^2; Hair, Hult, Ringle, & Sarstedt, 2017) are largely uninformative in the context of model selection as they always improve with model complexity. Relying solely on them may lead researchers to select an overly complex model, which overfits the data by tapping spurious patterns in that specific sample (Myung, 2000). As a result, the model will generalize poorly to other samples and will have a limited possibility of being replicated by other researchers. In contrast, parsimonious yet well-fitting models are more likely to be scientifically replicable, explainable, and exhibit higher predictive abilities (Bentler & Mooijaart, 1989; Shmueli & Koppius, 2011). Hence, empirical criteria used for selecting a model among alternatives must strike a balance between fit and parsimony (Myung, 2000), which none of the PLS criteria documented in textbooks (e.g., Hair, Hult, Ringle, & Sarstedt, 2017; Ramayah, Cheah, Chuah, & Memon, 2018), tutorial articles (e.g., Ali, Rasoolimanesh, Sarstedt, Ringle, & Ryu, 2018; Ringle, Sarstedt, Mitchell, & Gudergan, 2019), or recent research work on PLS (e.g., Franke & Sarstedt, 2019; Sarstedt, Ringle, & Hair, 2017) are designed to do. In addition, the use of structural path statistical significance (*p*-values) allows drawing statistical inference for judging theoretical support in a *selected* model, but offers no objective basis for selecting a model from a set of alternative models within the PLS framework (Aho, Derryberry & Peterson, 2014; Burnham & Anderson, 2002; Johnson & Omland, 2004; Raftery, 1995).

Addressing this critical gap in research, this study advocates empirically robust model selection practices by introducing criteria derived from information theory (e.g., Akaike, 1973) that allow identifying a parsimonious yet well-fitting model in a set of competing models within the PLS framework. These model selection criteria rely on the method of penalized-likelihood in which a term to penalize model complexity is added to the likelihood function (Sin & White, 1996). While they have a solid theoretical standing and are routinely used in econometrics and factor-based structural equation modeling (e.g., Chin & Todd, 1995; Sin & White, 1996; Zellner, Keuzenkamp, & McAleer, 2001), model selection criteria have not been adopted in PLS to date.[2]

In a series of three Monte Carlo studies, we compare and contrast the performance of model selection criteria with the PLS criteria under different scenarios that mimic the way prior literature has considered theoretically justified competing models, as evidenced in our review of empirical PLS studies published in a five-year period in four top IS journals. Our simulation results suggest that appropriate model selection cannot be achieved by relying on the PLS criteria (i.e., R^2, Adjusted R^2, GoF, and Q^2), as is the current practice in academic research. Instead, model selection criteria, in particular the Bayesian information criterion (BIC) and the Geweke-Meese criterion (GM), should be used due to their high model selection accuracy and ease of use. To support researchers in the adoption of these criteria, we introduce a five-step procedure that delineates the roles of model selection and statistical inference and discuss potential misconceptions that may arise when using the criteria. Our overarching aim in this paper is to encourage the practice of model selection in IS research—and more specifically in PLS-based studies—to foster the creation of generalizable theories.

2 Model Selection in PLS Studies: Current State of Affairs and Issues

2.1 Review of Current PLS Model Selection Practices in IS Research

To assess the degree to which IS researchers consider and compare theoretically motivated alternative models in their PLS analyses, we reviewed all articles published within a five-year period (April 2011-April 2016) in four top journals in the field: *Information Systems Research*, *Journal of the Association for Information Systems*, *Journal of Management Information Systems*, and *MIS Quarterly*. We conducted a full text search in the EBSCO Business Source Premier database using the keywords “partial least squares” and “PLS” to identify papers that performed PLS analyses. We also searched the online websites of the journals. Searching across the database and websites using the same keywords allowed us to verify that we had captured an exhaustive set of recent PLS articles in these journals. Because of our interest

[2] A potential reason is that PLS estimation does not rely on maximum likelihood. Yet, as we describe later, these criteria have been successfully implemented in a variety of nonlikelihood contexts, including least squares settings where their calculation is fairly straightforward, and can be readily adapted for use in PLS.

in PLS-based model comparison, nonempirical papers such as editorials, conceptual articles, and simulation studies were not considered (e.g., Aguirre-Urreta & Marakas, 2012; Becker, Rai, Ringle, & Völckner, 2013; Chin, Thatcher, & Wright, 2012).

This search resulted in 78 empirical articles that used PLS. In 34 of the 78 articles (43.59%), the authors considered multiple models in their PLS analyses. Appendix A presents more details about these 34 papers. In the majority of the 34 cases (25; 73.53%; Table A1 in Appendix A), however, researchers did not explicitly compare theoretically justified alternative models with the aim of selecting the best model from the set. Instead, authors specified multiple models to test the stability of effects by either adding or removing antecedent constructs and/or control variables (e.g., Arazy & Gellatly, 2012; Johnson & Cooper, 2015; Zhang, Venkatesh, & Brown, 2011), by comparing the original model with a saturated one (e.g., Armstrong, Brooks, & Riemenschneider, 2015), or by using alternative construct operationalizations (e.g., Grgecic, Holten, & Rosenkranz, 2015; Karahanna & Preston, 2013; Ray, Kim, & Morris, 2012). Several of these studies included moderator analyses (e.g., Sun, 2012; Venkatesh, Thong, & Xu, 2012; Xu, Benbasat, & Cenfetelli, 2011), which involve testing multiple models with and without interaction term(s) to test hypotheses related to the main and moderation effects (Hair, Hult, Ringle, & Sarstedt, 2017). In only 9 of the 34 studies (11.54%; Table A2 in Appendix A), the authors engaged in model comparisons in terms of Nuzzo (2015) by testing theoretically justified alternative model set-ups for explaining the phenomenon under study (e.g., Chandra, Srivastava, & Theng, 2012; Dinger, Thatcher, Treadway, Stepina, & Breland, 2015; Sykes, 2015). However, none of the studies attempted to empirically achieve a balance between model fit and parsimony.

2.2 Issues Related to Current PLS Model Selection Practices

2.2.1 Model Selection Based on Measures of Fit

Our literature review shows that researchers' most frequent justification for considering alternative models is to test the stability of effects by modifying the proposed model slightly, rather than generating alternative theoretical explanations. In the rare cases when they explored alternative model set-ups, researchers primarily compared the models in terms of the change in the models' R^2 values (8 of 9 studies), sometimes supplemented by an assessment of the changes in effect size (f^2). The justification for this choice may be that the model with the highest R^2 value does a better job than its rivals in capturing the signal in the data. While this choice may appear intuitive (or even forced due to the lack of model selection criteria in PLS), it can be problematic because the R^2, as a goodness-of-fit index, provides a single measure that includes both, (1) a fit to the signal, and (2) a fit to the noise. Thus, a model can produce high R^2 values due to a good fit to the noise despite its bad fit to the signal. A model's ability to fit the noise is closely correlated with its complexity and one can always improve the model's goodness-of-fit by increasing its complexity, such as by increasing the number of parameters (Myung & Pitt, 2004). In PLS, model complexity can increase due to a greater number of constructs, structural paths, or both (Hair, Sarstedt, & Ringle, 2019). Increased model complexity may give an impression of improved model fit, but it may come at the cost of its generalizability (Myung, 2000). For example, in the context of linear regression it is always possible to fit a complex *n*-degree polynomial through *n* data points exactly but that may not fit other data well (Forster, 2000).

Within the context of PLS, prior research has shown that R^2 and related statistics such as Tenenhaus, Amato, and Esposito Vinzi's (2004) goodness-of-fit index (GoF) are unsuitable for model selection. Specifically, Henseler and Sarstedt (2013) found that neither the GoF nor the relative GoF (GoF_{rel}) were able to separate valid models from invalid ones. The tendency of both R^2 and GoF-based measures to improve with model complexity means that these indices will almost always favor complex models over parsimonious ones, thus resulting in overfitting. Similarly, Evermann and Tate's (2010) simulation study shows that these and related criteria display a "bewildering range of behavior" (p. 16) while evaluating and assessing misspecified models. Rönkkö and Evermann (2013) report similar results.

2.2.2 Model Selection Based on the Null Hypothesis Paradigm

In one study in our review (Xue, Zhang, Ling, & Zhao, 2013), the authors compared models solely based on an assessment of the path coefficients and their significance (*p*-values). However, this approach to model selection raises issues of its own. First, the null hypothesis testing paradigm is limited by and becomes "strained" due to its two critical requirements: (1) the models being compared must be nested, and (2) one of the competing models must be assumed to be the "true" model (Aho, Derryberry, & Peterson, 2014). In practice, models under consideration may be nonnested and approximations of reality (i.e., not "true"). Second, the null hypothesis paradigm is structured in such a way that the null hypothesis cannot be directly *supported* by evidence; one can only fail to reject it in a given data set. More importantly, the null hypothesis paradigm cannot offer valid support for the alternative hypothesis that researchers truly seek

(Cohen, 1994). Hence, the null hypothesis exists only in a state of "suspended disbelief" (Wagenmakers, 2007). The American Statistical Association has recently cautioned that an overreliance on *p*-values, and especially the prevalent thresholds (e.g., 1%, 5% significance levels), may cause selective inference and lead researchers to "cherry-pick" promising findings, the so-called "*p*-value hacking" issue (Wasserstein & Lazar, 2016). This holds especially for PLS where statistical inference typically relies on bootstrapping, which requires researchers to choose specific settings (e.g., sign change options, bootstrap samples used). These choices, however, can have a significant bearing on the results (Rönkkö, McIntosh, & Antonakis, 2015). Third, the null hypothesis testing paradigm offers no guidance for uncovering well-fitting models that are also parsimonious, thereby ignoring the principle of Occam's razor. Fourth, the dependence of *p*-values on sample sizes may itself introduce biases (Lin, Lucas, & Shmueli, 2013). Finally, because the null hypothesis paradigm is the primary method of drawing inference and reporting results in PLS studies, its use as a model comparison tool can potentially induce publication bias (Easterbook et al., 1991; Rosenthal, 1979). Instead, the method for selecting a model should be kept separate from the method used for model inference and reporting. Raftery (1995) discusses in detail the problems associated with the use of *p*-values in model selection, including the practice of inclusion and exclusion of control variables.

Due to the reasons mentioned above, the practice of model selection based on R^2 measures and path significances in PLS studies is troublesome (Sobel, 2000). Instead, model selection should be driven not so much by the evaluation of particular hypotheses and maximized variance in a given setting (data), as is the current practice, but by a focus on the generalizability of the model by balancing model fit and complexity, also referred to as the "bias-variance tradeoff" (Wit, Heuvel, & Romeijn, 2012). A more complex model generalizes poorly to new data sets because it overfits the original data by absorbing random error (Myung 2000). In contrast, parsimonious models are likely to be generalizable and outperform complex models in their out-of-sample predictive abilities (Shmueli & Koppius, 2011).

Because generalizability cannot be estimated from a given sample, it is achieved by trading off model fit with complexity. In fact, generalizability is considered the "formal implementation of Occam's razor" (Myung & Pitt, 2004). In addition to its empirical role in predictive relevance and generalizability, parsimony is also regarded by many social scientists as an important ingredient in theory development (e.g., Gregor, 2006; Simon, 2001), precisely because it "explains much by little" (Friedman, 1994; p.153). Thus, generalizability-driven model selection has an integral role in the creation of "consilient theories", that is, theories that unify and systematize a field by explaining the facts taken from several domains (Thagard, 1978).

Model selection criteria strive to achieve the goal of maximizing generalizability by weighing goodness-of-fit relative to model complexity and allow the consideration of both nested and nonnested models without necessarily requiring that any of the models be "true" (Aho et al., 2014; Myung & Pitt, 2004). The next section provides more details about these criteria.

3 Model Selection Criteria

The research on developing methods to select the best model among a set of competing models has a distinguished history in the regression literature. The simplest criterion that might be considered in PLS is the R^2, which is calculated as:

$$R_k^2 = 1 - \frac{SS_{error}(k)}{SS_{total}} \qquad (1)$$

where $SS_{error}(k)$ is the sum of squared errors for the k^{th} model in a set of models and SS_{total} is total sum of squares. However, given that R^2 will increase as predictors are added to the model and hence will always select a more complex model, regression researchers have widely used the Adjusted R^2, which attempts to correct for model complexity. It is given by:

$$Adjusted\ R_k^2 = 1 - \left(\frac{n-1}{n-p_k}\right)\left(\frac{SS_{error}(k)}{SS_{total}}\right) \qquad (2)$$

where p_k is the number of coefficients (predictors plus intercept) in the k^{th} model. Taking the number of coefficients into account leads to a conceptual improvement over the R^2 because an effort is made to discount improvement in fit resulting solely from the model complexity. However, Berk (2008, p. 29) notes that "Adjusted R^2 lacks much formal justification" as the criterion is not based on rigorous statistical theory.

In the late 1960s and the early 1970s, model selection criteria that penalize model complexity in the interest of the principle of parsimony began to appear in the literature. These criteria were developed under the framework of information theory, a mathematical theory of communication, which studies the transmission, processing, extraction, and utilization of information (e.g., Akaike, 1973). McQuarrie and Tsai (1998) categorized the development of these criteria in two parallel streams of work. The first stream of model selection criteria seek to select a model that is *closest* to the *unknown* true model that generates the observed data and thereby defines the correlation patterns among the variables of interest. Relevant criteria in this stream include Akaike's (1970) final prediction error (FPE), Mallow's Cp (Mallows, 1973), Akaike's (1973) information criterion (AIC), Sugiura's (1978) corrected AIC (AICc), and McQuarrie and Tsai's (1998) unbiased AIC (AICu). The earliest selection criteria, FPE and Cp, both relied on the L_2 norm as the basis of measuring the distance between the true model and a candidate model (McQuarrie & Tsai, 1998), which is defined as:

$$\Delta L_2(M_t, M_a) = \left\| \mu_{M_t} - \mu_{M_a} \right\|^2, \quad (3)$$

where M_t is the true model with mean μ_{M_t}, and M_a is the candidate model with mean μ_{M_a}. In contrast to FPE and Cp, AIC, AICc, and AICu are based on the information theoretic notion of the Kullback-Leibler (KL) discrepancy to measure the distance, which is defined as:

$$\Delta KL(M_t, M_a) = E_{F_t}\left[\log \frac{f_t(x)}{f_a(x)}\right], (4)$$

where M_t is the true model with density $f_t(x)$ and distribution F_t and M_a is the candidate model. The main advantage of the AIC-type criteria is that they can be used to measure the *relative* distances of competing models from the unknown true model, even when the absolute distance to the true model is unknown. This characteristic allows researchers to compare the relative distances of several competing models and select the model closest to the unknown true model. This is equivalent to selecting the model with the smallest value on these criteria. All the distance-based criteria discussed here—FPE, Cp, AIC, AICc and AICu—are asymptotically efficient (McQuarrie & Tsai, 1998), which means that they tend to select the model with the minimum mean squared error between the unknown true model and a candidate model as the sample size increases (Shibata, 1980).

The goal of the second stream of model selection criteria is to provide an estimate of the posterior probability of a model being true and choose the model that maximizes this probability on a given data set. These model selection criteria are considered asymptotically consistent, which means that when the true model is included in the set of models being considered, the probability that the criteria will select it approaches unity with the increase in sample size (McQuarrie & Tsai, 1998; Hastie et al., 2009). Examples of such criteria include Schwarz's (1978) Bayesian information criterion (BIC), Geweke and Meese's (1981) criterion (GM), Hannan and Quinn's (1979) criterion (HQ) and McQuarrie and Tsai's (1998) corrected HQ criterion (HQc). Generally, researchers select a model with the smallest value on these criteria.

The model selection criteria from both the streams can be written as a function of the maximum value of the likelihood function. In models that are estimated using maximum likelihood (e.g., linear and logistic regression), the computation of the criteria is straightforward, and there exists extensive literature regarding their performance in various scenarios and methodological contexts, such as mixtures of normal distributions (e.g., Biernacki, Celeux, & Govaert, 2000; Bozdogan, 1994; Celeux & Soromenho, 1996), mixture regression models (e.g., Andrews & Currim, 2003a; Becker, Ringle, Sarstedt, & Völckner, 2015; Hawkins, Allen, & Stromberg, 2001), and mixture logit models (e.g., Andrews & Currim, 2003b).

Model comparison has also been proposed in the factor-based structural equation modeling context where "comparing the fit of alternative models has become a standard procedure" (Kumar & Sharma, 1999, p. 171). However, these comparisons typically involve testing nested models (Anderson & Gerbing, 1988) using the chi-squared difference test (Rust, Lee, & Valente, 1999). Comparing nonnested models is less common in factor-based structural equation modeling and requires different measures. Among these, the model selection criteria feature most prominently (e.g., Rust et al., 1999), and their performance has also been evaluated in several simulation studies. For example, Rust, Simester, Brodie, and Nilikant (1995) found that AIC and BIC perform better than computationally intensive jackknife or split sample-based criteria. Similarly, Homburg (1991) found these criteria to outperform simple cross-validation.

In contrast to factor-based structural equation modeling, PLS estimation is not performed using maximum likelihood but rather relies on nonlinear estimation by iterative least squares. In this scenario, closed-form formulas for the different model selection criteria do not exist. However, when the error distribution is normal with a constant variance, the maximum likelihood-based formulas can be written as a function of the model residuals, and specifically the sum of squared residuals (SS_{Error}) (Burnham & Anderson, 2002, p. 63; see also McQuarrie & Tsai, 1998). These least squares formulations using the model residuals are merely a special case of the equivalent likelihood estimation (Burnham & Anderson, 2002; Navarro & Myung, 2005). In fact, model selection criteria based on SS_{Error} are routinely computed in least squares settings (Anderson, Burnham, & Thompson, 2000; Burnham & Anderson, 2001; Burnham, Anderson, & Huyvaert, 2011; Johnson & Omland, 2004; Li, Morris, & Martin, 2002; Navarro & Myung, 2005; Symonds & Moussalli, 2010; Yamaoka, Nakagawa, & Uno, 1978), as well as in more general settings such as weighted least squares and even nonlinear models (e.g., Spiess & Neumeyer, 2010). In all these cases, the SS_{error} is used in place of the maximum likelihood value. We use these versions of the formulas in our study because SS_{Error} can be easily computed from PLS models.[3]

[3] While formal presentations of model selection criteria typically rely on maximum likelihood estimation, other general approaches outside the likelihood framework have also been developed (e.g., Konishi & Kitagawa, 1996; Pan,

Appendix B presents the details of all the criteria discussed in this section and explains their computation. As shown in the formulas in Table B1, each criterion can be written as a combination of two terms: the first term can be interpreted as a measure of lack of model fit, while the second term can be interpreted as the penalty for increasing model complexity (Burnham & Anderson, 2002). Therefore, these criteria try to achieve a trade-off between model fit and model complexity (Burnham & Anderson, 2001).

4 Monte Carlo Study

As the statistical properties of model selection criteria are largely beyond the reach of established asymptotic theory (e.g., Vrieze, 2012), Monte Carlo studies have become the norm for evaluating their performance (e.g., Andrews & Currim, 2002a, b; Becker et al., 2015; Hawkins et al., 2001). Following this standard practice, we conducted a Monte Carlo study under different model and data constellations. Specifically, our simulation study considers the following model selection criteria:

- Asymptotically efficient criteria: AIC, AICc, AICu, Cp, and FPE
- Asymptotically consistent criteria: BIC, GM, HQ, and HQc

As several prior studies in top IS journals have relied on the PLS criteria for comparing competing models (e.g., Plouffe et al., 2001; Venkatesh, Brown, Maruping, & Bala, 2008; Venkatesh et al., 2003; Venkatesh et al., 2012; Wixom & Todd, 2005), we include the R^2, Q^2, and GoF in our study. In addition, we include the Adjusted R^2 which is another PLS criterion that Sarstedt, Wilczynski, and Melewar (2013) brought forward in this context (Hair, Hult, Ringle, & Sarstedt, 2017).

4.1 Study Design

Drawing on the recommendations by Paxton, Curran, Bollen, Kirby, and Chen (2001), we utilize models of similar structure and complexity as those commonly encountered in IS research, such as the UTAUT model (Venkatesh et al., 2003; 2008) or other models of information systems success (e.g., Iyengar, Sweeney, & Montealegre, 2015; Park, Sharman, & Rao, 2015; Polites & Karahanna, 2012). Furthermore, the models are similar to those used in prior PLS-based simulation studies (e.g., Dijkstra & Henseler, 2015; Reinartz, Haenlein, & Henseler, 2009; Ringle, Sarstedt, & Schlittgen, 2014). All competing structural models in our study have five reflectively measured latent variables, three of which are exogenous (ξ_1, ξ_2 *and* ξ_3), while two are endogenous (η_1 and η_2).[4] Each construct has four items. The focus of our investigation is the target endogenous construct η_2.

We created three separate simulation set-ups to mimic possible model selection scenarios, depending on the breadth of variables available and the set of competing models the researcher is able to theorize. A core assumption in the three situations we discuss below is that the researcher generates a number of competing models, all of which are not incorrect, and at least one of which is a consistent (with correctly specified paths) but parsimonious version of reality—and this version is included in the set of competing models. This is a reasonable assumption in exploratory settings where the researcher is likely to possess only a partial or incomplete knowledge of reality due to evolving theory and may not have access to the true model that generates the observations. Thus, the researcher is able to correctly specify a *subset* of theoretical linkages in at least one of the competing models. This approach is consistent with the way that several studies have examined competing models that are theoretically justified. For example, in an effort to "further test the robustness of the proposed model and examine if the proposed configuration does explain the maximum variance in the final dependent variable", Chandra et al. (2012, p. 817) modified their originally hypothesized model by dropping two paths and comparing the resulting models' R^2 values with the original one. Similarly, Dinger et al. (2015) contrasted their original model against a reduced model, which only included those paths that the prior estimation rendered significant. Other studies have considered competing models by dropping constructs from the original model (e.g., Tan, Benbasat, & Cenfetelli, 2013) or including an additional construct (e.g., Sykes, 2015), which our final simulation set-up considers. In the following, we describe the three set-ups beginning with the least likely to occur in practice to the most likely:

Case 1: The first case mimics the condition where the researcher has access to all latent variables that helped create the observed data. In addition, it assumes that the researcher has correctly theorized and included the true model (i.e., the data-generating model) that depicts the reality in its entirety in the competing model set-up. We note that this condition is very

2001a; b; Yafune, Funatogawa, & Ishiguro, 2005). Model selection methods are now available for nonparametric regression, splines, kernel methods, martingales, generalized estimating equations (Burnham & Anderson, 2002), and even genetic algorithms (Pond, Posada, Gravenor, Woelk, & Frost, 2006).

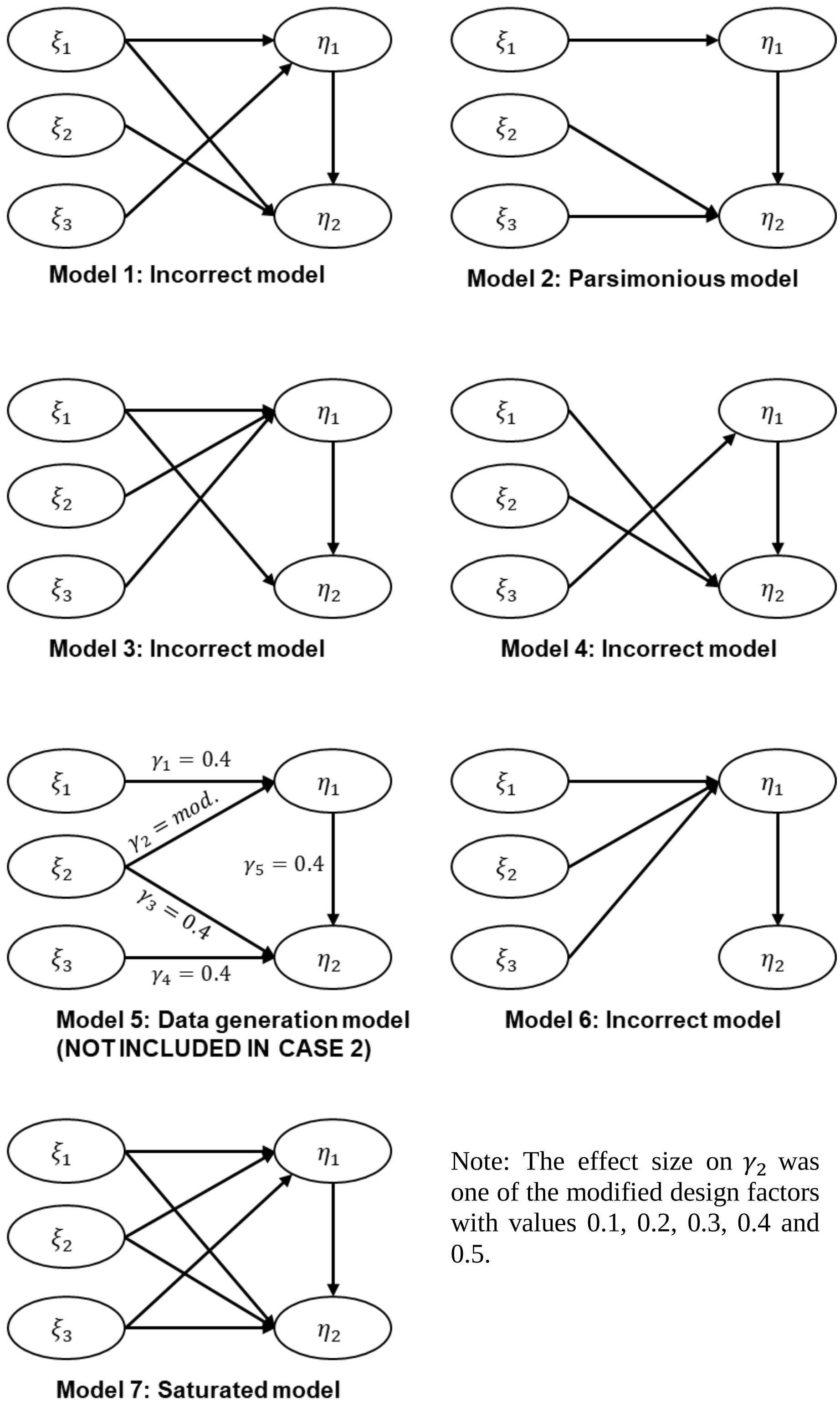
[4] We focus on structural model comparisons for two reasons. First, the structural paths in a PLS model typically represent the theorized hypotheses under consideration, and second, the structure of the measurement model should be established prior to any analysis of the structural relations.

unlikely to occur in practice because exploratory research typically involves situations where researchers are unlikely to have access to or awareness of the complete set of variables and linkages that formed the observed reality. This is especially true in IS research where context-specific variables play significant roles in affecting reality (e.g., Addas, 2010; Moon & Kim, 2001; Raymond, 1985).

This condition consists of a set of seven competing structural models (Figure 1), which may be thought of as representing a set of competing hypotheses that the researcher wants to explore. Here, Model 5 depicts the data generation process (i.e., the true model) and is included in the competing set. Models 1, 3, 4, and 6 are incorrectly specified; that is, they have incorrect linkages that are not theoretically consistent with the data generation process. Model 2 is a parsimonious and consistent version of the data generation process; that is, it includes only correctly specified paths, although not all. Finally, Model 7 is a fully saturated model with all possible paths explaining η_2 that serves as the basis for assessing overall fit and represents the common situation in research where R^2 acts as the sole basis for model comparisons. The goal of model selection in this set-up is to reject incorrect models on the basis that they represent incorrect theoretical linkages. Furthermore, the saturated model should also be rejected in the interest of parsimony (Occam's razor) and due to the inclusion of incorrect linkages. Since the data generating process (Model 5) is included in the competing set-up, its selection represents the best-case scenario for the researcher. Failing that, the next best option is the selection of Model 2, which represents a parsimonious but consistent version of the data generation process.

Case 2: This condition mimics the case where the researcher has access to all latent variables that helped create the observed data but fails to incorporate all relevant theoretical linkages among them; that is, the researcher does not include the data generating model in the competing model set-up. Thus, this condition is similar to Case 1 with one critical difference: there are six competing models and the researcher has failed to theorize and include the data-generating model in the competing set. That is, Model 5 has *not* been included (Figure 1). In this case, the ideal model selection criterion should choose Model 2, which is the parsimonious and consistent version of the data generation process, and rule out other models. We note that, similar to Case 1, this condition is also unlikely to occur in practice because researchers typically do not have access to the exhaustive set of constructs that generated the data.

Case 3: The third, and more realistic, condition assumes the existence of a latent variable that is unavailable to the researcher at the time of data collection or otherwise unknown, but that affected the observed data. There are several reasons why a researcher may not possess certain latent variables that took part in the data generation process. These include, but are not limited to, lack of access to data or unavailability, weak or nascent theory, theory exploration, incorrect logic, or accidental neglect. As seen in Figure 2, the data-generating model (Model X) is not included in the competing set and has an extraneous variable (ξ_4) that is not available to the researcher but that nevertheless took part in the data generation process. We think it is more likely that researchers will identify and collect variables with stronger effects based on logic and theory but may ignore variables with weaker effects. Thus, the extraneous variable (ξ_4) has a weak direct effect (-0.1) on η_2 that makes it more prone to omission. The goal of model selection in this context is to rule out incorrect and saturated model specifications, and select either Models 2 or 5, both of which now represent parsimonious but consistent explanations of the data generation process.



Note: The effect size on γ_2 was one of the modified design factors with values 0.1, 0.2, 0.3, 0.4 and 0.5.

Figure 1. Simulation Models (Cases 1 and 2)

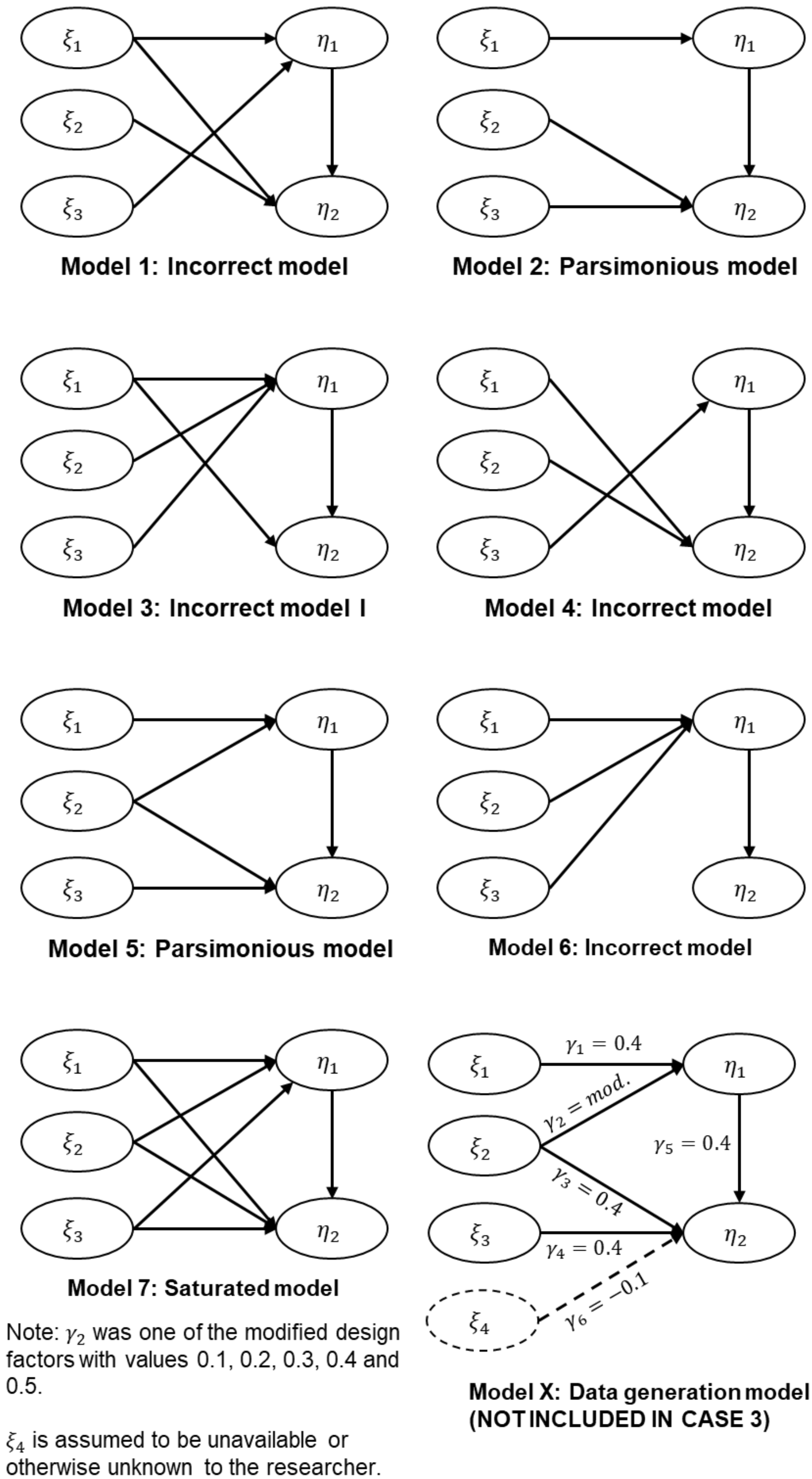


Figure 2. Simulation Models (Case 3)

4.2 Experimental Conditions and Data Generation Method

In each of the three cases outlined above, data were generated for the underlying data-generating model (i.e., Model 5 in Cases 1 and 2, Model X in Case 3) by manipulating the following experimental conditions:

- Six conditions of sample size (50, 100, 150, 200, 250, and 500);[5]
- Five conditions of effect size on the path $\xi_1 \rightarrow \eta_2$ (0.1, 0.2, 0.3, 0.4, and 0.5);
- Three factor loading patterns with different levels of average variance extracted (AVE):
    - High AVE with loadings: (0.9, 0.9, 0.9, and 0.9),
    - Moderate AVE with loadings: (0.8, 0.8, 0.8, and 0.8), and
    - Low AVE with loadings: (0.7, 0.7, 0.7, and 0.7).

The data generation used in this study draws on a procedure similar to the one that Schlittgen (2015) presents in his SEGIRLS package for the R statistical software. The approach generates data from the covariance matrix of the indicators, followed by a Cholesky decomposition and multiplication with a sample of the sought data distribution.[6] The simulation study considers the case of normally distributed data as recent research has shown that PLS estimates are consistent when estimating data from composite model populations (Sarstedt, Hair, Ringle et al., 2016; Hair, Hult, Ringle, Sarstedt, & Thiele, 2017)—as it is the case in our research.[7] All simulations were run in the R computing cluster environment (R Development Core Team, 2014) using the *sempls* (Monecke, 2012) and the *snowfall* package for parallel computing (Knaus, 2013). Drawing on Reinartz et al. (2009), we ran 300 replications for each of the 90 simulation conditions, yielding a total of 27,000 cases. For each case, we used the corresponding data set as the input to estimate each of the models under consideration (e.g., Models 1-7 in Case 1).

The dependent variable of interest was a binary vector with a value of 1 denoting the model selected by the criterion in that run, and 0 otherwise. In the case of the PLS criteria (R^2, Adjusted R^2, GoF, and Q^2) a model was selected if the criterion achieved the highest value among the competing models. In contrast, in the case of asymptotically efficient and consistent model selection criteria, a model was selected if the criterion achieved the lowest value among the competing models (McQuarrie & Tsai, 1998).

5 Analysis and Results

5.1 Overall Model Selection Rates

We first describe the overall model selection rates (i.e., the percentage of times the criteria selected each model in the set) across the three cases. Recall that Case 1 represents the (unlikely) scenario in which both the full data generation (Model 5) and parsimonious but consistent (Model 2) models are included in the competing set. In this case, which has 7 competing models, a completely random choice would select each model 14.3% of the time, on average. Table 1A presents the overall model-specific selection rates across all experimental conditions for this case. For example, the first cell in Table 1A indicates that across all the simulation runs, AICc chose Model 1 in 7% of the cases.

In the following, we use the term “success” to denote the choice of Model 2 or 5. Furthermore, we denote the R^2, Adjusted R^2, GoF, and Q^2 as PLS criteria, as these criteria have previously been used in the PLS context.

In terms of the PLS criteria, GoF performed the worst (overall success rate 4.07%) due to its overwhelming tendency to favor the saturated model. GoF selected Models 2 and 5 in only 0.07% and 4% of cases, respectively. R^2 gave somewhat better performance (overall success rate 30.53%) but had a strong preference for the saturated model (69.13%). Q^2 had a better success rate (50.66%) but it tended to select an incorrect model (Model 6) with some frequency (19.97%). Adjusted R^2 performed better than other PLS criteria (overall success rate 69.31%), because of the penalty it applies for overfitting. However, it still selected the saturated model in 30.14% of the runs, suggesting that it did not sufficiently penalize overfitting. Overall, none of the PLS criteria performed satisfactorily.

[5] Note that sample sizes of 50 are generally unacceptable when using PLS. However, in line with prior PLS-based simulation studies (Aguirre-Urreta & Rönkkö, 2018; Goodhue, Lewis, & Thompson, 2012; Rönkkö & Evermann, 2013), we considered this factor level to explore the criteria’s performance under such a boundary condition.

[6] For a more detailed explanation on the data generation approach, see Ringle et al. (2014).

[7] Nevertheless, we also tested the models using non-normal data (chi-squared distributed with $df = 3$, t-distributed with $df = 5$, and uniform). The results were highly stable across different data distributions.

Table 1A. Overall Model Selection Rates (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

Criteria		Model # 1	2	3	4	5	6	7	Success rate
Asymptotically efficient	AICc	0.07	69.23	0.02	0.64	15.03	0.09	14.93	84.26
	AICu	0.06	73.36	0.04	0.65	16.18	0.25	9.47	89.54
	AIC	0.06	67.93	0.02	0.6	14.67	0.04	16.68	82.61
	CP	0.07	73.01	0.03	0.63	16.09	0.08	10.11	89.09
	FPE	0.06	67.96	0.02	0.6	14.68	0.04	16.65	82.63
Asymptotically consistent	BIC	0.06	77.85	0.04	0.65	17.49	0.53	3.38	95.34
	GM	0.07	79.34	0.07	0.66	18.02	0.82	1.02	97.37
	HQc	0.06	74.93	0.03	0.66	16.64	0.3	7.38	91.57
	HQ	0.06	74.13	0.03	0.64	16.35	0.17	8.63	90.47
PLS	GoF	0.05	0.07	0.01	0.01	4	0.03	95.87	4.07
	R^2	0.01	28.44	0	0.32	2.1	0.03	69.13	30.53
	Adj R^2	0.04	57.5	0	0.5	11.81	0.02	30.14	69.31
	Q^2	3.67	25.41	8.34	2.8	25.25	19.97	14.72	50.66

Note: Success rate denotes the choice of Model 2 or 5.

The set of asymptotically efficient criteria (AIC, AICc, AICu, FPE, and Cp) had significantly better model selection rates than the PLS criteria. AICu and Cp had success rates approaching 90%, followed by AICc (84.26%), FPE (82.63%), and AIC (82.61%). While these criteria soundly rejected incorrect model specifications, they had a fair tendency to select the saturated model (overall rates range from 9.47% to 16.68%), suggesting that their penalty functions, while much better than Adjusted R^2, still left some doubts about their utility. Finally, these criteria displayed stronger preference for Model 2 (67.93% to 73.36%) over Model 5 (14.67% to 16.18%).

The asymptotically consistent criteria (BIC, GM, HQ, and HQc) performed the best among all criteria. In particular, BIC and GM had almost perfect success rates (95.34% and 97.37%, respectively) across all experimental conditions, followed by HQc and HQ (91.57% and 90.47%, respectively). The penalty functions employed by BIC and GM helped reject the saturated model in almost all the cases (Model 7 selection rates: 3.38% and 1.02%, respectively). BIC and GM also displayed stronger preference for Model 2 (77.85% and 79.34%, respectively) over Model 5 (17.49% and 18.02%, respectively) due to the strong penalty functions. These results suggest that even in the unlikely scenario where the reality has been captured (as Model 5) and included in the competing set, researchers might not have a high chance of selecting it. However, they can be very confident that the use of these criteria, in particular BIC and GM, will lead them to select a consistent (but parsimonious) version of reality, which could mean the possibility of missing one or more paths but not having incorrect paths or paths that are irrelevant (as in the saturated model). This is a not an unreasonable compromise in exploratory research.

Case 2 represents the scenario where the data-generating model (i.e., Model 5) was left out of the competing set even when all relevant variables were accessible to the researcher. This situation may occur due to theoretical oversight in exploratory research. In this case, which includes six competing models, a completely random choice would select each model 16.7% of the time, on average. Table 1B presents the overall model selection rates across all experimental conditions for this case. The trends in the results are very similar to Case 1, and in the absence of Model 5, all the criteria selected Model 2 with high regularity. BIC and GM again emerged as the best performing criteria with overall success rates of 95.30% and 97.35%, respectively.

Table 1B. Overall Model Selection Rates (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included

Criteria		Model # 1	2	3	4	6	7	Success rate
Asymptotically efficient	AICc	0.07	84.10	0.02	0.64	0.09	15.09	84.10
	AICu	0.06	89.47	0.04	0.65	0.25	9.54	89.47
	AIC	0.06	82.37	0.02	0.60	0.04	16.92	82.37
	CP	0.07	88.97	0.03	0.63	0.08	10.23	88.97
	FPE	0.06	82.40	0.02	0.60	0.04	16.89	82.40
Asymptotically consistent	BIC	0.06	95.30	0.04	0.65	0.54	3.42	95.30
	GM	0.07	97.35	0.07	0.66	0.82	1.04	97.35
	HQc	0.06	91.51	0.03	0.66	0.30	7.44	91.51
	HQ	0.06	90.40	0.03	0.64	0.17	8.70	90.40
PLS	GoF	0.06	0.21	0.03	0.01	0.03	99.71	0.21
	R^2	0.01	29.76	0.00	0.32	0.03	69.91	29.76
	Adj R^2	0.04	68.97	0.00	0.50	0.02	30.48	68.97
	Q^2	3.67	49.01	8.35	2.85	20.01	16.25	49.01

Note: Success rate denotes the choice of Model 2.

The success rates for asymptotically efficient criteria ranged between 82.4% and 89.47%. In contrast, results for the PLS criteria largely parallel those from Case 1. GoF and R^2 achieve low success rates and have a strong tendency to select the saturated model. Adjusted R^2 and Q^2 perform somewhat better than in Case 1 with success rates of 68.97% and 49.01%, respectively. Yet, both criteria have a pronounced tendency of selecting the saturated model. These results strongly suggest that the model selection criteria, in particular BIC and GM, can help researchers rule out incorrect specifications and select theoretically consistent models in contexts where theoretical oversight is likely.

Case 3 represents the more likely and realistic scenario where some relevant variables are unavailable to the researcher and render the data-generating model out of reach. In this case, the best a researcher could hope for is the ability to select a consistent but parsimonious version of reality (i.e., either Model 2 or 5). In this case, which includes seven competing models, a completely random choice would select each model 14.3% of the time, on average. Table 1C presents the overall model selection rates across all experimental conditions for this case.

The results in Table 1C largely parallel those in previous cases. Again, GoF and R^2 performed the worst (success rates of 3.79% and 30.48%, respectively) among all criteria and displayed significant tendencies to favor the saturated model. Q^2 and Adjusted R^2 performed better but their performances were not satisfactory (50.27% and 69.53%, respectively). The asymptotically consistent criteria performed the best, and BIC and GM, in particular, displayed near perfect success rates (95.22% and 97.35%, respectively), while the asymptotically efficient criteria plateaued around 82-90%. Thus, the use of consistent criteria, such as BIC and GM, may allow selecting consistent but parsimonious versions of reality with near certainty, even in cases where reality consists of variables that are out of reach. Next, to gain deeper insights about how the experimental conditions affected the model selection rates, we present the analysis broken down by specific experimental conditions.

Table 1C. Overall Model Selection Rates (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included

Criteria		Model #							Success rate
		1	2	3	4	5	6	7	
Asymptotically efficient	AICc	0.06	69.24	0.03	0.64	14.85	0.08	15.1	84.1
	AICu	0.06	73.29	0.05	0.64	16.01	0.19	9.76	89.3
	AIC	0.05	67.93	0.02	0.58	14.47	0.06	16.9	82.4
	CP	0.06	72.98	0.03	0.65	15.93	0.07	10.29	88.91
	FPE	0.05	67.95	0.02	0.58	14.47	0.06	16.87	82.43
Asymptotically consistent	BIC	0.06	78.06	0.06	0.63	17.17	0.5	3.52	95.22
	GM	0.07	79.76	0.09	0.68	17.58	0.79	1.03	97.35
	HQc	0.07	75.03	0.06	0.64	16.44	0.21	7.55	91.47
	HQ	0.06	74.03	0.04	0.64	16.2	0.12	8.91	90.23
PLS	GoF	0.03	0.05	0	0.01	3.74	0.01	96.17	3.79
	R^2	0.01	28.27	0	0.35	2.21	0.01	69.16	30.48
	Adj R^2	0.04	57.77	0.01	0.5	11.76	0.01	29.91	69.53
	Q^2	3.3	25.63	8.14	2.93	24.64	20.45	15.11	50.27

Note: Success rate denotes the choice of Model 2 or 5.

5.2 Model Selection Rates by Sample Size

Our simulation design considers six sample size conditions (50, 100, 150, 200, 250, and 500). We expected that, in general, model selection rates would improve with sample size due to a gain in statistical power. However, the results tell a more nuanced story where performances of some criteria improved more than others, while others remained stagnant, and in some cases deteriorated. Ideally, we would like to pinpoint criteria that not only provide good performance at smaller sample sizes, but that also show significant improvements with an increase in sample size.

The overall trends in the results are very similar across Cases 1, 2, and 3. Tables 2A, 2B, and 2C present detailed model specific success rates broken down by sample size. In terms of the PLS criteria, the performances of GoF and R^2 slightly deteriorate with an increase in sample size across all three cases. For example, the overall success rates of R^2 decreases from 32.58%, at sample size 50, to 27.76%, at sample size 500, for the most likely scenario (Case 3, Table 2C). On closer inspection, this performance decrease can be attributed to an increase in the preference for the saturated model with increasing sample size. More precisely, R^2 favored the saturated model (Model 7) 65.38% times at sample size 50, but this increased to 72.24% at sample size 500. In contrast, the Q^2 success rate improved from 39.18% to 57.51% but was still low. This pattern persisted across all three cases. These results strongly suggest that in the case of PLS criteria, the narrative "the greater the sample size, the higher the accuracy" is incorrect as far as model selection is concerned. The very same holds for Adjusted R^2, whose success rates show a similar development as those of GoF and R^2 for varying sample sizes.

Table 2A. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

Sample size	Criteria	Model #							Success rate
		1	2	3	4	5	6	7	
50	AICc	0.40	68.49	0.13	3.36	15.92	0.53	11.21	84.40
	AICu	0.33	69.62	0.22	3.40	16.40	1.51	8.53	86.03
	AIC	0.36	65.53	0.11	3.18	15.00	0.22	15.64	80.54
	CP	0.38	70.32	0.16	3.33	16.50	0.47	8.87	86.82
	FPE	0.36	65.64	0.11	3.18	15.03	0.22	15.51	80.67
	BIC	0.33	70.51	0.22	3.38	16.85	3.13	5.59	87.36
	GM	0.36	71.97	0.40	3.45	17.52	4.74	1.58	89.49
	HQc	0.33	70.22	0.20	3.42	16.72	1.82	7.31	86.94
	HQ	0.31	68.84	0.20	3.33	15.94	1.04	10.36	84.78
	GoF	0.29	0.29	0.07	0.07	5.85	0.16	93.54	6.13
	R^2	0.07	30.78	0.00	1.73	3.27	0.16	64.17	34.04
	Adj R^2	0.24	57.51	0.02	2.67	12.09	0.11	27.43	69.60
	Q^2	9.24	20.78	11.20	7.82	18.18	20.44	12.96	38.96
100	AICc	0.02	70.33	0.00	0.44	15.04	0.00	14.16	85.38
	AICu	0.02	74.36	0.00	0.49	16.16	0.00	8.98	90.51
	AIC	0.02	68.62	0.00	0.42	14.76	0.00	16.18	83.38
	CP	0.02	73.73	0.00	0.44	16.00	0.00	9.80	89.73
	FPE	0.02	68.62	0.00	0.42	14.76	0.00	16.18	83.38
	BIC	0.02	78.51	0.00	0.49	17.42	0.07	3.49	95.93
	GM	0.07	80.24	0.00	0.47	17.93	0.18	1.11	98.18
	HQc	0.02	75.62	0.00	0.49	16.53	0.00	7.33	92.16
	HQ	0.02	74.36	0.00	0.49	16.16	0.00	8.98	90.51
	GoF	0.00	0.07	0.00	0.00	4.36	0.00	95.58	4.42
	R^2	0.02	28.91	0.00	0.18	2.38	0.00	68.51	31.29
	Adj R^2	0.02	58.16	0.00	0.29	12.20	0.00	29.33	70.36
	Q^2	5.73	23.93	10.02	4.11	23.67	19.07	13.73	47.60
150	AICc	0.00	72.02	0.00	0.00	14.56	0.00	13.42	86.58
	AICu	0.00	76.04	0.00	0.00	15.71	0.00	8.24	91.76
	AIC	0.00	70.84	0.00	0.00	14.22	0.00	14.93	85.07
	CP	0.00	75.58	0.00	0.00	15.67	0.00	8.76	91.24
	FPE	0.00	70.87	0.00	0.00	14.22	0.00	14.91	85.09
	BIC	0.00	80.16	0.00	0.00	16.82	0.00	3.02	96.98
	GM	0.00	81.67	0.00	0.02	17.13	0.00	1.18	98.80
	HQc	0.00	77.36	0.00	0.00	16.11	0.00	6.53	93.47

Table 2A. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

	HQ	0.00	76.38	0.00	0.00	15.87	0.00	7.76	92.24
	GoF	0.00	0.04	0.00	0.00	3.20	0.00	96.76	3.24
	R^2	0.00	29.13	0.00	0.00	2.20	0.00	68.67	31.33
	Adj R^2	0.00	60.00	0.00	0.00	11.44	0.00	28.56	71.44
	Q^2	3.02	27.22	8.38	2.20	24.38	21.11	13.69	51.60
200	**AICc**	0.00	69.02	0.00	0.02	15.42	0.00	15.53	84.44
	AICu	0.00	74.00	0.00	0.02	16.64	0.00	9.33	90.64
	AIC	0.00	68.33	0.00	0.02	15.18	0.00	16.47	83.51
	CP	0.00	73.42	0.00	0.02	16.69	0.00	9.87	90.11
	FPE	0.00	68.33	0.00	0.02	15.18	0.00	16.47	83.51
	BIC	0.00	78.76	0.00	0.02	18.02	0.00	3.20	96.78
	GM	0.00	80.47	0.00	0.02	18.62	0.00	0.89	99.09
	HQc	0.00	75.31	0.00	0.02	17.09	0.00	7.58	92.40
	HQ	0.00	74.71	0.00	0.02	16.87	0.00	8.40	91.58
	GoF	0.00	0.02	0.00	0.00	3.69	0.00	96.29	3.71
	R^2	0.00	28.27	0.00	0.02	1.96	0.00	69.76	30.22
	Adj R^2	0.00	57.16	0.00	0.02	12.24	0.00	30.58	69.40
	Q^2	2.42	26.31	7.58	1.69	27.62	19.96	14.47	53.93
250	**AICc**	0.00	69.42	0.00	0.00	14.29	0.00	16.29	83.71
	AICu	0.00	74.27	0.00	0.00	15.67	0.00	10.07	89.93
	AIC	0.00	68.58	0.00	0.00	14.00	0.00	17.42	82.58
	CP	0.00	74.13	0.00	0.00	15.53	0.00	10.33	89.67
	FPE	0.00	68.58	0.00	0.00	14.00	0.00	17.42	82.58
	BIC	0.00	80.27	0.00	0.00	17.00	0.00	2.73	97.27
	GM	0.00	81.56	0.00	0.00	17.64	0.00	0.80	99.20
	HQc	0.00	76.38	0.00	0.00	16.16	0.00	7.47	92.53
	HQ	0.00	75.96	0.00	0.00	16.11	0.00	7.93	92.07
	GoF	0.00	0.00	0.00	0.00	3.98	0.00	96.02	3.98
	R^2	0.00	28.22	0.00	0.00	1.71	0.00	70.07	29.93
	Adj R^2	0.00	57.76	0.00	0.00	11.42	0.00	30.82	69.18
	Q^2	1.20	27.56	7.02	0.78	27.36	20.60	15.49	54.91
500	**AICc**	0.00	66.07	0.00	0.00	14.98	0.00	18.96	81.04
	AICu	0.00	71.87	0.00	0.00	16.49	0.00	11.64	88.36
	AIC	0.00	65.69	0.00	0.00	14.89	0.00	19.42	80.58
	CP	0.00	70.84	0.00	0.00	16.13	0.00	13.02	86.98
	FPE	0.00	65.69	0.00	0.00	14.89	0.00	19.42	80.58
	BIC	0.00	78.89	0.00	0.00	18.84	0.00	2.27	97.73

Table 2A. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

	GM	0.00	80.16	0.00	0.00	19.29	0.00	0.56	99.44
	HQc	0.00	74.69	0.00	0.00	17.22	0.00	8.09	91.91
	HQ	0.00	74.53	0.00	0.00	17.13	0.00	8.33	91.67
	GoF	0.00	0.00	0.00	0.00	2.96	0.00	97.04	2.96
	R^2	0.00	25.31	0.00	0.00	1.07	0.00	73.62	26.38
	Adj R^2	0.00	54.40	0.00	0.00	11.47	0.00	34.13	65.87
	Q^2	0.38	26.64	5.84	0.22	30.29	18.62	18.00	56.93

Note: Success rate denotes the choice of Model 2 or 5.

Table 2B. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included.

Sample size	Criteria	Model #						Success rate
		1	2	3	4	6	7	
50	**AICc**	0.40	84.13	0.13	3.36	0.56	11.45	84.13
	AICu	0.33	85.93	0.22	3.40	1.51	8.62	85.93
	AIC	0.36	80.20	0.11	3.18	0.22	15.98	80.20
	CP	0.38	86.64	0.16	3.33	0.47	9.05	86.64
	FPE	0.36	80.36	0.11	3.18	0.22	15.82	80.36
	BIC	0.33	87.27	0.22	3.38	3.16	5.66	87.27
	GM	0.36	89.46	0.40	3.45	4.74	1.60	89.46
	HQc	0.33	86.91	0.20	3.42	1.82	7.33	86.91
	HQ	0.31	84.60	0.20	3.33	1.04	10.54	84.60
	GoF	0.33	0.69	0.16	0.07	0.16	98.86	0.69
	R^2	0.07	33.02	0.00	1.73	0.16	65.20	33.02
	Adj R^2	0.24	69.24	0.02	2.67	0.11	27.78	69.24
	Q^2	9.29	37.00	11.24	8.09	20.69	14.20	37.00
100	**AICc**	0.02	85.31	0.00	0.44	0.00	14.22	85.31
	AICu	0.02	90.51	0.00	0.49	0.00	8.98	90.51
	AIC	0.02	83.22	0.00	0.42	0.00	16.33	83.22
	CP	0.02	89.69	0.00	0.44	0.00	9.84	89.69
	FPE	0.02	83.22	0.00	0.42	0.00	16.33	83.22
	BIC	0.02	95.87	0.00	0.49	0.07	3.56	95.87
	GM	0.07	98.18	0.00	0.47	0.18	1.11	98.18
	HQc	0.02	92.07	0.00	0.49	0.00	7.42	92.07
	HQ	0.02	90.51	0.00	0.49	0.00	8.98	90.51
	GoF	0.00	0.20	0.00	0.00	0.00	99.80	0.20
	R^2	0.02	30.51	0.00	0.18	0.00	69.29	30.51

Table 2B. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included.

	Adj R^2	0.02	70.11	0.00	0.29	0.00	29.58	70.11
	Q^2	5.73	46.07	10.02	4.11	19.07	15.27	46.07
150	**AICc**	0.00	86.38	0.00	0.00	0.00	13.62	86.38
	AICu	0.00	91.67	0.00	0.00	0.00	8.33	91.67
	AIC	0.00	84.69	0.00	0.00	0.00	15.31	84.69
	CP	0.00	91.07	0.00	0.00	0.00	8.93	91.07
	FPE	0.00	84.71	0.00	0.00	0.00	15.29	84.71
	BIC	0.00	96.96	0.00	0.00	0.00	3.04	96.96
	GM	0.00	98.78	0.00	0.02	0.00	1.20	98.78
	HQc	0.00	93.42	0.00	0.00	0.00	6.58	93.42
	HQ	0.00	92.16	0.00	0.00	0.00	7.84	92.16
	GoF	0.00	0.11	0.00	0.00	0.00	99.89	0.11
	R^2	0.00	30.47	0.00	0.00	0.00	69.53	30.47
	Adj R^2	0.00	71.18	0.00	0.00	0.00	28.82	71.18
	Q^2	3.02	50.31	8.38	2.20	21.11	14.98	50.31
200	**AICc**	0.00	84.27	0.00	0.02	0.00	15.71	84.27
	AICu	0.00	90.56	0.00	0.02	0.00	9.42	90.56
	AIC	0.00	83.38	0.00	0.02	0.00	16.60	83.38
	CP	0.00	90.09	0.00	0.02	0.00	9.89	90.09
	FPE	0.00	83.38	0.00	0.02	0.00	16.60	83.38
	BIC	0.00	96.76	0.00	0.02	0.00	3.22	96.76
	GM	0.00	99.07	0.00	0.02	0.00	0.91	99.07
	HQc	0.00	92.29	0.00	0.02	0.00	7.69	92.29
	HQ	0.00	91.56	0.00	0.02	0.00	8.42	91.56
	GoF	0.00	0.11	0.00	0.00	0.00	99.89	0.11
	R^2	0.00	29.44	0.00	0.02	0.00	70.53	29.44
	Adj R^2	0.00	69.00	0.00	0.02	0.00	30.98	69.00
	Q^2	2.42	52.11	7.58	1.69	19.96	16.29	52.11
250	**AICc**	0.00	83.56	0.00	0.00	0.00	16.44	83.56
	AICu	0.00	89.87	0.00	0.00	0.00	10.13	89.87
	AIC	0.00	82.29	0.00	0.00	0.00	17.71	82.29
	CP	0.00	89.51	0.00	0.00	0.00	10.49	89.51
	FPE	0.00	82.29	0.00	0.00	0.00	17.71	82.29
	BIC	0.00	97.27	0.00	0.00	0.00	2.73	97.27
	GM	0.00	99.18	0.00	0.00	0.00	0.82	99.18
	HQc	0.00	92.51	0.00	0.00	0.00	7.49	92.51
	HQ	0.00	92.00	0.00	0.00	0.00	8.00	92.00

Table 2B. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included.

	GoF	0.00	0.13	0.00	0.00	0.00	99.87	0.13	
	R^2	0.00	29.24	0.00	0.00	0.00	70.76	29.24	
	Adj R^2	0.00	68.80	0.00	0.00	0.00	31.20	68.80	
	Q^2	1.20	53.13	7.02	0.78	20.60	17.27	53.13	
500	**AICc**	0.00	80.93	0.00	0.00	0.00	19.07	80.93	
	AICu	0.00	88.27	0.00	0.00	0.00	11.73	88.27	
	AIC	0.00	80.42	0.00	0.00	0.00	19.58	80.42	
	CP	0.00	86.84	0.00	0.00	0.00	13.16	86.84	
	FPE	0.00	80.42	0.00	0.00	0.00	19.58	80.42	
	BIC	0.00	97.67	0.00	0.00	0.00	2.33	97.67	
	GM	0.00	99.42	0.00	0.00	0.00	0.58	99.42	
	HQc	0.00	91.87	0.00	0.00	0.00	8.13	91.87	
	HQ	0.00	91.58	0.00	0.00	0.00	8.42	91.58	
	GoF	0.00	0.02	0.00	0.00	0.00	99.98	0.02	
	R^2	0.00	25.84	0.00	0.00	0.00	74.16	25.84	
	Adj R^2	0.00	65.49	0.00	0.00	0.00	34.51	65.49	
	Q^2	0.38	55.42	5.84	0.22	18.62	19.51	55.42	

Note: Success rate denotes the choice of Model 2.

Table 2C. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included.

Sample size	**Criteria**	**Model #**							**Success rate**
		1	**2**	**3**	**4**	**5**	**6**	**7**	
50	**AICc**	0.31	67.31	0.18	3.60	15.96	0.49	12.17	83.27
	AICu	0.36	68.89	0.31	3.58	16.51	1.11	9.25	85.40
	AIC	0.29	64.51	0.11	3.24	15.02	0.36	16.49	79.53
	CP	0.33	69.81	0.18	3.65	16.68	0.40	8.97	86.48
	FPE	0.29	64.56	0.11	3.27	15.02	0.36	16.42	79.58
	BIC	0.36	70.16	0.36	3.47	17.00	2.80	5.87	87.16
	GM	0.38	71.92	0.51	3.62	17.43	4.54	1.60	89.35
	HQc	0.38	70.11	0.38	3.60	16.80	1.24	7.50	86.91
	HQ	0.31	67.78	0.24	3.62	16.20	0.71	11.15	83.98
	GoF	0.20	0.22	0.02	0.04	6.13	0.04	93.44	6.36
	R^2	0.07	29.09	0.02	1.96	3.49	0.07	65.38	32.58
	Adj R^2	0.24	56.00	0.04	2.82	12.60	0.09	28.23	68.60
	Q^2	8.38	21.58	12.24	7.18	17.60	19.69	14.22	39.18
100	**AICc**	0.02	70.31	0.00	0.16	15.16	0.00	14.36	85.47

Table 2C. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included.

	AICu	0.02	73.98	0.00	0.16	16.20	0.02	9.62	90.18
	AIC	0.02	68.47	0.00	0.16	14.76	0.00	16.60	83.22
	CP	0.02	73.53	0.00	0.16	16.09	0.00	10.20	89.62
	FPE	0.02	68.56	0.00	0.16	14.76	0.00	16.51	83.31
	BIC	0.02	78.36	0.00	0.22	17.29	0.18	3.93	95.64
	GM	0.02	80.33	0.00	0.27	17.87	0.20	1.31	98.20
	HQc	0.02	75.33	0.00	0.18	16.60	0.02	7.84	91.93
	HQ	0.02	73.98	0.00	0.16	16.22	0.02	9.60	90.20
	GoF	0.00	0.04	0.00	0.00	3.78	0.00	96.18	3.82
	R^2	0.00	29.18	0.00	0.11	2.73	0.00	67.98	31.91
	Adj R^2	0.02	58.62	0.00	0.13	12.07	0.00	29.16	70.69
	Q^2	5.02	25.27	9.62	4.20	21.33	20.87	13.93	46.60
150	**AICc**	0.00	69.62	0.00	0.07	15.56	0.00	14.76	85.18
	AICu	0.00	73.84	0.00	0.09	16.82	0.00	9.24	90.67
	AIC	0.00	68.82	0.00	0.07	15.20	0.00	15.91	84.02
	CP	0.00	73.53	0.00	0.09	16.76	0.00	9.62	90.29
	FPE	0.00	68.82	0.00	0.07	15.20	0.00	15.91	84.02
	BIC	0.00	78.76	0.00	0.11	17.84	0.02	3.27	96.60
	GM	0.00	80.58	0.00	0.18	18.33	0.00	0.91	98.91
	HQc	0.00	75.38	0.00	0.09	17.02	0.00	7.51	92.40
	HQ	0.00	74.53	0.00	0.09	16.91	0.00	8.47	91.44
	GoF	0.00	0.02	0.00	0.00	3.13	0.00	96.84	3.16
	R^2	0.00	29.04	0.00	0.02	2.20	0.00	68.73	31.24
	Adj R^2	0.00	58.60	0.00	0.02	12.40	0.00	28.98	71.00
	Q^2	3.18	25.22	7.64	2.58	25.38	21.93	14.09	50.60
200	**AICc**	0.00	69.93	0.00	0.00	14.58	0.00	15.49	84.51
	AICu	0.00	74.67	0.00	0.00	15.69	0.00	9.64	90.36
	AIC	0.00	68.73	0.00	0.00	14.40	0.00	16.87	83.13
	CP	0.00	74.09	0.00	0.00	15.60	0.00	10.31	89.69
	FPE	0.00	68.73	0.00	0.00	14.40	0.00	16.87	83.13
	BIC	0.00	80.22	0.00	0.00	17.09	0.00	2.69	97.31
	GM	0.00	81.80	0.00	0.00	17.53	0.00	0.67	99.33
	HQc	0.00	76.51	0.00	0.00	16.13	0.00	7.36	92.64
	HQ	0.00	75.89	0.00	0.00	16.02	0.00	8.09	91.91
	GoF	0.00	0.00	0.00	0.00	2.96	0.00	97.04	2.96
	R^2	0.00	28.13	0.00	0.00	1.76	0.00	70.11	29.89
	Adj R^2	0.00	58.20	0.00	0.00	11.69	0.00	30.11	69.89

Table 2C. Model Selection Rates Broken Down by Sample Size (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included.

	Q^2	1.78	26.42	7.13	1.89	26.29	21.09	15.40	52.71
250	**AICc**	0.00	70.69	0.00	0.00	13.80	0.00	15.51	84.49
	AICu	0.00	75.44	0.00	0.00	15.09	0.00	9.47	90.53
	AIC	0.00	69.76	0.00	0.00	13.49	0.00	16.76	83.24
	CP	0.00	74.58	0.00	0.00	14.93	0.00	10.49	89.51
	FPE	0.00	69.78	0.00	0.00	13.49	0.00	16.73	83.27
	BIC	0.00	80.71	0.00	0.00	16.44	0.00	2.84	97.16
	GM	0.00	82.27	0.00	0.00	16.71	0.00	1.02	98.98
	HQc	0.00	77.16	0.00	0.00	15.64	0.00	7.20	92.80
	HQ	0.00	76.69	0.00	0.00	15.51	0.00	7.80	92.20
	GoF	0.00	0.02	0.00	0.00	2.96	0.00	97.02	2.98
	R^2	0.00	27.91	0.00	0.00	1.60	0.00	70.49	29.51
	Adj R^2	0.00	58.98	0.00	0.00	10.78	0.00	30.24	69.76
	Q^2	1.27	27.56	6.38	1.56	27.47	20.02	15.80	55.02
500	**AICc**	0.00	67.60	0.00	0.00	14.07	0.00	18.33	81.67
	AICu	0.00	72.93	0.00	0.00	15.76	0.00	11.31	88.69
	AIC	0.00	67.27	0.00	0.00	13.98	0.00	18.76	81.24
	CP	0.00	72.33	0.00	0.00	15.51	0.00	12.16	87.84
	FPE	0.00	67.27	0.00	0.00	13.98	0.00	18.76	81.24
	BIC	0.00	80.13	0.00	0.00	17.33	0.00	2.53	97.47
	GM	0.00	81.69	0.00	0.00	17.62	0.00	0.69	99.31
	HQc	0.00	75.67	0.00	0.00	16.44	0.00	7.89	92.11
	HQ	0.00	75.31	0.00	0.00	16.36	0.00	8.33	91.67
	GoF	0.00	0.00	0.00	0.00	3.49	0.00	96.51	3.49
	R^2	0.00	26.24	0.00	0.00	1.51	0.00	72.24	27.76
	Adj R^2	0.00	56.22	0.00	0.00	11.02	0.00	32.76	67.24
	Q^2	0.18	27.73	5.82	0.18	29.78	19.09	17.22	57.51

Note: Success rate denotes the choice of Model 2 or 5.

The asymptotically efficient criteria (AIC, AICc, AICu, Cp, and FPE) had more nuanced performances that peaked at sample size 150 and decreased thereafter across all three cases. For example, in Case 3, the performance of AIC at sample sizes 50, 150, and 500 varied from 79.53% to 84.02% to 81.24%, respectively. Similar to R^2, this decrease after sample size 150 was fueled by greater preference for the saturated model, again suggesting that the penalty functions employed by these criteria were not strong enough to counter the gain in explained variance with increasing sample size. This observation limits the utility of these criteria.

Recall that criteria that correctly select the data-generating model with an increase in the sample size are considered asymptotically consistent (McQuarrie & Tsai, 1998). The performances of the consistent criteria (BIC, GM, HQ, and HQc) improved considerably with an increase in the sample size, suggesting that the penalty employed by these criteria guarded against the tendency to select the saturated model. BIC and GM, in particular, provided robust performances and improved their (Case 3) success rates from 87.16% and 89.35%, at sample size 50, to near perfect 97.47% and 99.31%, at sample size 500, respectively. In fact, their success rates crossed 95% and 98% at the relatively low sample size of 100, which points to their practical utility in exploratory research. Thus, PLS researchers can be very confident that utilizing these criteria can allow them to select consistent models with high probability.

5.3 Model Selection Rates by Effect Size

Based on our model set-ups for all three cases, we expected that all selection criteria would increasingly favor Model 5 over Model 2 as the $\xi_1 \rightarrow \eta_2$ path strength increased. In general, the results corroborate this premise but with a few exceptions. Since the pattern of results was similar in all three cases (Tables 3A, 3B, and 3C), we take Case 3 as the exemplar for our discussion (Table 3C).

The overall success rates of the PLS criteria deteriorated significantly with an increase in the $\xi_1 \rightarrow \eta_2$ path strength. For example, at effect size 0.1, R^2, Adjusted R^2, and Q^2 had overall success rates of 37.30%, 71.15%, and 67.39%, respectively. This reduced to 21.24%, 66.35%, and 30.48% at effect size 0.5. The reduction in performance of R^2 and Adjusted R^2 can be directly attributed to their increased preference for the saturated model (Model 7). This means that while both R^2 and Adjusted R^2 reduced their preference for Model 2, the subsequent increase in their preference for Model 5 was small and overshadowed by their stronger predisposition for the saturated model. The Q^2 success rate also dropped with an increase in effect size but mainly due to preference for an incorrect model (Model 6). Since Q^2 is a predictive metric, it is difficult to judge whether the choice is correct or incorrect in terms of predictive power, because wrong models can sometimes predict better than correct ones (Shmueli, 2010). While Model 6 may have been a better predictive model at larger effect sizes, it was clearly not consistent with the data- generating model.

The success rates of the efficient criteria (AIC, AICc, AICu, FPE, and Cp) also deteriorated, although to a lesser extent than those of the PLS criteria, due to their increased tendency to favor the saturated model with an increase in effect size. For example, at effect size 0.1, AIC's preference for Models 2, 5, and 7 was 76.54%, 7.13%, and 15.74%, respectively. At effect size 0.5, this preference evolved to 55.43%, 24.91%, and 18.67%, respectively. Thus, while these criteria did increase their preference for Model 5 over Model 2 as we had expected, their performance still left a lot to be desired.

The consistent criteria (BIC, GM, HQ, and HQc) again performed best among all criteria. BIC and GM hardly showed any deterioration in performance with a change in effect size. At effect size 0.1, BIC's preference for Models 2, 5, and 7 was 87.15%, 8.35%, and 3.44%, respectively, with an overall success rate (i.e., preference for Model 2 *or* 5) of 95.50%; GM's preferences were 88.94%, 8.56%, and 1.20% with an overall success rate of 97.50%. At effect size 0.5, this evolved for BIC to 65.31%, 29.50%, and 3.56%, respectively, with an overall success rate of 94.81%; For GM, this evolved to 66.88%, 30.13%, and 1.04%, respectively, with an overall success rate of 97.02%. This means that as the effect size on the $\xi_1 \rightarrow \eta_2$ path increased, BIC and GM improved their preference for Model 5 over Model 2 as expected, while strongly avoiding their tendency to select the saturated model.

Table 3A. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

Effect size	Criteria	Model #							Success rate
		1	2	3	4	5	6	7	
0.1	AICc	0.04	78.72	0.04	0.44	7.39	0.04	13.34	86.11
	AICu	0.04	82.87	0.06	0.43	7.81	0.19	8.61	90.69
	AIC	0.04	77.46	0.02	0.44	7.26	0.02	14.76	84.72
	CP	0.04	82.22	0.04	0.43	7.69	0.06	9.54	89.90
	FPE	0.04	77.50	0.02	0.44	7.26	0.02	14.72	84.76
	BIC	0.04	87.59	0.04	0.46	8.41	0.39	3.08	96.00
	GM	0.06	89.00	0.06	0.48	8.63	0.65	1.13	97.63
	HQc	0.04	84.39	0.06	0.44	8.02	0.20	6.85	92.41
	HQ	0.04	83.44	0.06	0.44	7.91	0.11	8.00	91.35
	GoF	0.00	0.02	0.00	0.00	4.26	0.00	95.74	4.28
	R^2	0.00	36.70	0.00	0.19	0.98	0.00	62.14	37.69
	Adj R^2	0.04	66.48	0.00	0.33	5.69	0.00	27.47	72.17
	Q^2	1.48	32.52	3.54	1.65	35.24	6.26	19.35	67.76
0.2	AICc	0.09	76.44	0.02	0.72	8.93	0.11	13.69	85.37
	AICu	0.07	80.85	0.02	0.76	9.70	0.20	8.39	90.56
	AIC	0.07	75.13	0.02	0.69	8.69	0.06	15.35	83.81
	CP	0.07	80.46	0.02	0.74	9.56	0.09	9.06	90.02
	FPE	0.07	75.17	0.02	0.69	8.69	0.06	15.32	83.85
	BIC	0.07	85.19	0.02	0.72	10.57	0.37	3.06	95.76
	GM	0.07	86.80	0.06	0.59	10.93	0.59	0.96	97.72
	HQc	0.07	82.63	0.00	0.76	9.96	0.26	6.32	92.59
	HQ	0.07	81.67	0.04	0.76	9.72	0.17	7.58	91.39
	GoF	0.02	0.00	0.04	0.00	4.30	0.02	95.65	4.30
	R^2	0.04	34.35	0.00	0.33	1.19	0.00	64.10	35.54
	Adj R^2	0.04	64.06	0.00	0.50	6.93	0.02	28.47	70.98
	Q^2	2.04	30.43	5.89	1.61	30.56	12.15	17.44	60.98
0.3	AICc	0.04	70.06	0.02	0.48	13.80	0.11	15.50	83.86
	AICu	0.06	74.59	0.02	0.50	14.86	0.26	9.72	89.45
	AIC	0.04	68.63	0.02	0.43	13.43	0.02	17.45	82.06
	CP	0.04	74.44	0.02	0.44	14.97	0.02	10.08	89.41
	FPE	0.04	68.67	0.02	0.43	13.43	0.02	17.41	82.10
	BIC	0.06	79.46	0.02	0.52	16.02	0.61	3.32	95.49
	GM	0.06	80.91	0.00	0.57	16.58	1.00	0.89	97.49
	HQc	0.06	76.17	0.02	0.50	15.34	0.31	7.61	91.50
	HQ	0.04	75.35	0.02	0.46	15.10	0.20	8.84	90.45

Table 3A. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

	GoF	0.02	0.04	0.00	0.02	3.96	0.00	95.98	4.00
	R^2	0.00	29.89	0.00	0.22	1.57	0.00	68.33	31.46
	Adj R^2	0.04	57.91	0.02	0.33	10.87	0.00	30.84	68.78
	Q^2	3.06	25.78	8.52	2.50	25.96	18.59	15.78	51.74
0.4	**AICc**	0.13	64.80	0.00	0.65	18.93	0.09	15.42	83.72
	AICu	0.09	69.02	0.02	0.70	20.33	0.31	9.53	89.35
	AIC	0.13	63.57	0.00	0.65	18.57	0.04	17.05	82.15
	CP	0.13	68.88	0.02	0.67	20.30	0.11	9.90	89.18
	FPE	0.13	63.57	0.00	0.65	18.59	0.04	17.03	82.17
	BIC	0.09	73.20	0.02	0.67	21.76	0.67	3.60	94.96
	GM	0.11	75.01	0.07	0.70	22.37	0.83	0.89	97.39
	HQc	0.09	70.35	0.02	0.69	20.83	0.39	7.64	91.19
	HQ	0.09	69.78	0.02	0.65	20.52	0.19	8.77	90.30
	GoF	0.07	0.15	0.02	0.02	3.85	0.06	95.92	4.00
	R^2	0.02	23.17	0.00	0.39	2.83	0.07	73.58	26.00
	Adj R^2	0.09	53.46	0.00	0.61	14.96	0.04	30.86	68.43
	Q^2	4.83	21.59	10.63	3.33	20.02	27.80	11.98	41.61
0.5	**AICc**	0.06	56.11	0.04	0.89	26.13	0.09	16.69	82.24
	AICu	0.04	59.46	0.07	0.87	28.19	0.30	11.08	87.65
	AIC	0.04	54.87	0.04	0.81	25.43	0.06	18.77	80.30
	CP	0.06	59.02	0.04	0.89	27.93	0.11	11.97	86.94
	FPE	0.04	54.87	0.04	0.81	25.43	0.06	18.77	80.30
	BIC	0.04	63.80	0.09	0.87	30.70	0.63	3.87	94.50
	GM	0.06	65.00	0.15	0.94	31.61	1.02	1.22	96.61
	HQc	0.04	61.11	0.07	0.89	29.04	0.35	8.51	90.15
	HQ	0.04	60.41	0.04	0.89	28.48	0.20	9.95	88.89
	GoF	0.13	0.15	0.00	0.02	3.65	0.06	96.07	3.80
	R^2	0.02	18.07	0.00	0.48	3.91	0.06	77.51	21.98
	Adj R^2	0.02	45.57	0.00	0.70	20.61	0.04	33.07	66.19
	Q^2	6.93	16.72	13.13	4.93	14.46	35.04	9.06	31.19

Note: Success rate denotes the choice of Model 2 or 5.

Table 3B. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included

Effect size	Criteria	Model #						Success rate
		1	2	3	4	6	7	
0.1	AICc	0.04	86.02	0.04	0.44	0.04	13.43	86.02
	AICu	0.04	90.63	0.06	0.43	0.19	8.67	90.63
	AIC	0.04	84.56	0.02	0.44	0.02	14.93	84.56
	CP	0.04	89.85	0.04	0.43	0.06	9.59	89.85
	FPE	0.04	84.59	0.02	0.44	0.02	14.89	84.59
	BIC	0.04	96.00	0.04	0.46	0.39	3.08	96.00
	GM	0.06	97.63	0.06	0.48	0.65	1.13	97.63
	HQc	0.04	92.39	0.06	0.44	0.20	6.87	92.39
	HQ	0.04	91.30	0.06	0.44	0.11	8.06	91.30
	GoF	0.00	0.02	0.02	0.00	0.00	99.98	0.02
	R^2	0.00	37.35	0.00	0.19	0.00	62.47	37.35
	Adj R^2	0.04	71.94	0.00	0.33	0.00	27.69	71.94
	Q^2	1.48	65.81	3.54	1.67	6.28	21.26	65.81
0.2	AICc	0.09	85.22	0.02	0.72	0.11	13.83	85.22
	AICu	0.07	90.52	0.02	0.76	0.20	8.43	90.52
	AIC	0.07	83.61	0.02	0.69	0.06	15.56	83.61
	CP	0.07	89.96	0.02	0.74	0.09	9.11	89.96
	FPE	0.07	83.65	0.02	0.69	0.06	15.52	83.65
	BIC	0.07	95.70	0.02	0.72	0.37	3.11	95.70
	GM	0.07	97.69	0.06	0.59	0.59	1.00	97.69
	HQc	0.07	92.56	0.00	0.76	0.26	6.35	92.56
	HQ	0.07	91.35	0.04	0.76	0.17	7.61	91.35
	GoF	0.02	0.02	0.04	0.00	0.02	99.93	0.02
	R^2	0.04	35.02	0.00	0.33	0.00	64.62	35.02
	Adj R^2	0.04	70.80	0.00	0.50	0.02	28.65	70.80
	Q^2	2.04	58.94	5.91	1.61	12.19	19.41	58.94
0.3	AICc	0.04	83.69	0.02	0.48	0.11	15.67	83.69
	AICu	0.06	89.37	0.02	0.50	0.26	9.80	89.37
	AIC	0.04	81.85	0.02	0.43	0.02	17.65	81.85
	CP	0.04	89.24	0.02	0.44	0.02	10.24	89.24
	FPE	0.04	81.89	0.02	0.43	0.02	17.61	81.89
	BIC	0.06	95.46	0.02	0.52	0.61	3.33	95.46
	GM	0.06	97.46	0.00	0.57	1.00	0.91	97.46
	HQc	0.06	91.46	0.02	0.50	0.31	7.65	91.46
	HQ	0.04	90.43	0.02	0.46	0.20	8.85	90.43

Table 3B. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included

	GoF	0.04	0.07	0.00	0.02	0.00	99.89	0.07
	R^2	0.00	30.74	0.00	0.22	0.00	69.05	30.74
	Adj R^2	0.04	68.43	0.02	0.33	0.00	31.19	68.43
	Q^2	3.06	49.93	8.52	2.54	18.65	17.50	49.93
0.4	**AICc**	0.13	83.56	0.00	0.65	0.11	15.57	83.56
	AICu	0.09	89.28	0.02	0.70	0.31	9.60	89.28
	AIC	0.13	81.78	0.00	0.65	0.04	17.42	81.78
	CP	0.13	89.09	0.02	0.67	0.11	9.99	89.09
	FPE	0.13	81.80	0.00	0.65	0.04	17.40	81.80
	BIC	0.09	94.91	0.02	0.67	0.67	3.65	94.91
	GM	0.11	97.37	0.07	0.70	0.83	0.91	97.37
	HQc	0.09	91.09	0.02	0.69	0.39	7.73	91.09
	HQ	0.09	90.19	0.02	0.65	0.19	8.88	90.19
	GoF	0.07	0.30	0.07	0.02	0.06	99.57	0.30
	R^2	0.02	25.06	0.00	0.39	0.07	74.53	25.06
	Adj R^2	0.09	68.07	0.00	0.61	0.04	31.21	68.07
	Q^2	4.85	40.31	10.63	3.41	27.83	13.11	40.31
0.5	**AICc**	0.06	82.00	0.04	0.89	0.09	16.93	82.00
	AICu	0.04	87.54	0.07	0.87	0.30	11.19	87.54
	AIC	0.04	80.04	0.04	0.81	0.06	19.03	80.04
	CP	0.06	86.72	0.04	0.89	0.11	12.19	86.72
	FPE	0.04	80.06	0.04	0.81	0.06	19.01	80.06
	BIC	0.04	94.41	0.09	0.87	0.65	3.95	94.41
	GM	0.06	96.59	0.15	0.94	1.02	1.24	96.59
	HQc	0.04	90.06	0.07	0.89	0.35	8.60	90.06
	HQ	0.04	88.74	0.04	0.89	0.20	10.10	88.74
	GoF	0.15	0.65	0.00	0.02	0.06	99.20	0.65
	R^2	0.02	20.61	0.00	0.48	0.06	78.88	20.61
	Adj R^2	0.02	65.61	0.00	0.70	0.04	33.65	65.61
	Q^2	6.94	30.04	13.15	5.02	35.09	9.98	30.04

Note: Success rate denotes the choice of Model 2

Table 3C. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included

Effect size	Criteria	Model # 1	2	3	4	5	6	7	Success Rate
0.1	AICc	0.02	77.94	0.00	0.56	7.28	0.07	14.13	85.22
	AICu	0.02	82.31	0.00	0.56	7.80	0.15	9.17	90.11
	AIC	0.02	76.54	0.00	0.54	7.13	0.04	15.74	83.67
	CP	0.02	81.63	0.00	0.56	7.63	0.02	10.15	89.26
	FPE	0.02	76.59	0.00	0.54	7.13	0.04	15.69	83.72
	BIC	0.00	87.15	0.02	0.54	8.35	0.50	3.44	95.50
	GM	0.00	88.94	0.04	0.61	8.56	0.65	1.20	97.50
	HQc	0.02	84.09	0.04	0.57	8.02	0.15	7.11	92.11
	HQ	0.02	82.89	0.00	0.56	7.85	0.13	8.56	90.74
	GoF	0.00	0.00	0.00	0.00	4.48	0.00	95.52	4.48
	R^2	0.02	36.28	0.00	0.28	1.02	0.00	62.41	37.30
	Adj R^2	0.02	65.57	0.00	0.43	5.57	0.00	28.41	71.15
	Q^2	1.65	32.98	3.31	1.59	34.41	6.78	19.50	67.39
0.2	AICc	0.04	76.09	0.02	0.37	9.72	0.07	13.69	85.81
	AICu	0.06	79.63	0.04	0.37	10.46	0.19	9.26	90.09
	AIC	0.04	74.83	0.00	0.33	9.44	0.06	15.30	84.28
	CP	0.04	79.48	0.02	0.41	10.37	0.07	9.61	89.85
	FPE	0.04	74.83	0.00	0.33	9.44	0.06	15.30	84.28
	BIC	0.06	84.31	0.04	0.37	11.04	0.54	3.65	95.35
	GM	0.06	86.26	0.06	0.41	11.35	0.78	1.09	97.61
	HQc	0.06	81.52	0.04	0.35	10.65	0.22	7.17	92.17
	HQ	0.04	80.59	0.04	0.37	10.50	0.13	8.33	91.09
	GoF	0.02	0.00	0.00	0.00	3.94	0.02	96.04	3.94
	R^2	0.02	33.72	0.00	0.26	1.59	0.02	64.40	35.31
	Adj R^2	0.04	63.70	0.00	0.30	7.81	0.02	28.13	71.52
	Q^2	2.06	30.07	5.74	2.02	31.46	11.59	17.24	61.54
0.3	AICc	0.09	70.61	0.04	0.67	13.56	0.09	14.95	84.17
	AICu	0.09	74.41	0.07	0.67	14.74	0.20	9.82	89.15
	AIC	0.09	69.09	0.04	0.59	13.22	0.07	16.90	82.31
	CP	0.09	74.40	0.06	0.69	14.85	0.09	9.82	89.26
	FPE	0.09	69.11	0.04	0.61	13.22	0.07	16.86	82.33
	BIC	0.11	79.43	0.07	0.63	16.09	0.44	3.22	95.52
	GM	0.11	80.88	0.13	0.61	16.37	0.95	0.95	97.26
	HQc	0.11	76.15	0.07	0.67	15.26	0.22	7.52	91.41
	HQ	0.09	75.20	0.06	0.67	14.89	0.11	8.99	90.09

Table 3C. Model Selection Rates Broken Down by Effect Size (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included

	GoF	0.04	0.02	0.02	0.02	3.89	0.00	96.05	3.91
	R^2	0.02	29.61	0.00	0.37	2.37	0.02	67.63	31.98
	Adj R^2	0.07	59.11	0.02	0.52	11.00	0.04	29.25	70.11
	Q^2	2.74	26.20	8.65	2.33	25.06	20.31	14.91	51.26
0.4	**AICc**	0.06	65.06	0.04	0.67	18.15	0.07	15.96	83.20
	AICu	0.07	69.24	0.06	0.69	19.52	0.20	10.22	88.76
	AIC	0.04	63.74	0.02	0.61	17.67	0.06	17.87	81.41
	CP	0.06	69.07	0.04	0.70	19.39	0.06	10.69	88.46
	FPE	0.04	63.78	0.02	0.61	17.67	0.06	17.83	81.44
	BIC	0.07	74.07	0.06	0.69	20.85	0.52	3.74	94.93
	GM	0.09	75.85	0.06	0.72	21.50	0.89	0.89	97.35
	HQc	0.07	71.06	0.06	0.72	20.07	0.24	7.78	91.13
	HQ	0.06	69.91	0.06	0.69	19.81	0.11	9.37	89.72
	GoF	0.00	0.13	0.00	0.00	3.24	0.00	96.63	3.37
	R^2	0.00	24.20	0.00	0.39	2.37	0.00	73.04	26.57
	Adj R^2	0.02	54.28	0.00	0.54	14.24	0.00	30.93	68.52
	Q^2	3.89	21.91	10.59	3.43	18.78	28.59	13.09	40.69
0.5	**AICc**	0.07	56.52	0.06	0.93	25.56	0.09	16.78	82.07
	AICu	0.07	60.87	0.09	0.91	27.54	0.20	10.32	88.41
	AIC	0.07	55.43	0.04	0.81	24.91	0.07	18.67	80.33
	CP	0.09	60.31	0.04	0.89	27.39	0.09	11.19	87.70
	FPE	0.07	55.44	0.04	0.81	24.91	0.07	18.65	80.35
	BIC	0.07	65.31	0.11	0.94	29.50	0.50	3.56	94.81
	GM	0.07	66.88	0.15	1.04	30.13	0.69	1.04	97.02
	HQc	0.07	62.31	0.11	0.91	28.20	0.22	8.17	90.52
	HQ	0.07	61.56	0.06	0.94	27.96	0.13	9.28	89.52
	GoF	0.11	0.11	0.00	0.02	3.15	0.02	96.63	3.26
	R^2	0.00	17.52	0.02	0.44	3.72	0.02	78.30	21.24
	Adj R^2	0.07	46.19	0.02	0.70	20.17	0.02	32.84	66.35
	Q^2	6.17	16.98	12.41	5.28	13.50	34.96	10.81	30.48

Note: Success rate denotes the choice of Model 2 or 5.

5.4 Model Selection Rates by Loading Condition (AVE)

Each construct in our set-up had four items whose loadings were all set to either 0.7 (low AVE condition), 0.8 (moderate), or 0.9 (high AVE condition). We expected that an increase in item loadings and the resulting increase in AVE would allow for higher success rates for all criteria. Again, the pattern of results was similar across all three cases (Tables 4A, 4B, and 4C) and we use Case 3 (Table 4C) as the exemplar for subsequent discussion.

Contrary to our expectation, the overall success rate of R^2 significantly deteriorated with an increase in item loadings (36.94% for low AVE to 20.87% in high AVE). This decrease can again be attributed to R^2's tendency to favor the saturated model. The overall success rate and model preferences of Adjusted R^2 did not change appreciably. At the same time, we noticed a dramatic improvement in the success rate of Q^2, which increased from 34.69% in low AVE condition to 69.50% in high AVE, but was still not competitive with the efficient and consistent model selection criteria. In the low AVE condition, Q^2 selected incorrect models with fair frequency (e.g., Models 3 and 6 were selected 11.34% and 31.76% times, respectively). However, this improved appreciably in the high AVE condition (3.66% and 5.99%, respectively).

The performance of the asymptotically efficient criteria (AIC, AICc, AICu, FPE, and Cp) saw marginal improvements at higher AVE levels. This improvement was partly due to a reduction in the selection rate of Model 7. For example, the overall success rate of AIC increased marginally from 81.74% (low AVE) to 83.27% (high AVE) as its preference for the saturated model reduced from 17.12% to 16.59%.

The consistent criteria (BIC, GM, HQ, and HQc) also showed similar small improvements in their success rates at higher AVE levels. However, as compared to the efficient criteria, the performance of the consistent criteria started at a much higher base level. For example, GM's success rate improved from 96.17% (low AVE) to a near perfect 99.02% (high AVE), while BIC's success rate improved from 94.19% to 96.63%. Overall, the change in loading conditions had a small but positive impact on overall success rates of consistent criteria. This result underlines the utility of consistent criteria, in particular of BIC and GM, in exploratory research where the measurement models may be underdeveloped or evolving.

Table 4A. Model Selection Rates Broken Down by Loading (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

Loading condition	Criteria	Model #							Success rate
		1	2	3	4	5	6	7	
0.7	AICc	0.12	70.10	0.04	1.17	12.71	0.13	15.73	82.81
	AICu	0.09	74.38	0.07	1.20	13.86	0.42	10.00	88.23
	AIC	0.10	68.84	0.03	1.14	12.42	0.04	17.42	81.27
	CP	0.11	73.16	0.04	1.18	13.51	0.11	11.89	86.68
	FPE	0.10	68.86	0.03	1.14	12.42	0.04	17.41	81.28
	BIC	0.09	78.79	0.07	1.19	15.10	0.93	3.84	93.89
	GM	0.12	80.32	0.09	1.18	15.73	1.30	1.26	96.05
	HQc	0.09	75.74	0.06	1.21	14.31	0.53	8.06	90.06
	HQ	0.09	75.16	0.07	1.19	14.04	0.30	9.16	89.20
	GoF	0.09	0.11	0.03	0.02	5.42	0.04	94.35	5.53
	R^2	0.02	34.03	0.00	0.64	2.32	0.06	62.97	36.36
	Adj R^2	0.07	58.92	0.01	0.96	9.88	0.03	30.15	68.80
	Q^2	7.10	17.09	11.83	5.24	17.28	30.56	11.01	34.37
0.8	AICc	0.04	69.21	0.02	0.64	15.21	0.10	14.77	84.42
	AICu	0.04	73.07	0.04	0.63	16.36	0.27	9.59	89.42
	AIC	0.04	68.02	0.02	0.60	14.72	0.06	16.54	82.74
	CP	0.04	72.52	0.03	0.63	16.15	0.10	10.52	88.67

Table 4A. Model Selection Rates Broken Down by Loading (Percentages)
Case 1: All Variables Included; Data-Generating Model Included

	FPE	0.04	68.03	0.02	0.60	14.73	0.06	16.51	82.77
	BIC	0.04	77.42	0.04	0.61	17.76	0.53	3.59	95.18
	GM	0.04	78.95	0.08	0.63	18.25	0.82	1.22	97.20
	HQc	0.04	74.68	0.04	0.63	16.80	0.29	7.51	91.48
	HQ	0.03	73.84	0.03	0.64	16.50	0.17	8.78	90.34
	GoF	0.06	0.09	0.00	0.00	4.57	0.01	95.30	4.66
	R^2	0.02	31.10	0.00	0.31	2.31	0.00	66.27	33.41
	Adj R^2	0.03	58.18	0.00	0.49	11.93	0.01	29.36	70.11
	Q^2	3.20	23.74	9.41	2.43	23.76	23.68	13.92	47.50
0.9	**AICc**	0.04	68.37	0.00	0.10	17.18	0.03	14.28	85.55
	AICu	0.04	72.63	0.00	0.12	18.32	0.07	8.81	90.96
	AIC	0.04	66.93	0.00	0.07	16.88	0.01	16.07	83.81
	CP	0.04	73.33	0.00	0.09	18.60	0.02	7.91	91.94
	FPE	0.04	66.98	0.00	0.07	16.88	0.01	16.03	83.86
	BIC	0.04	77.33	0.00	0.14	19.62	0.13	2.72	96.96
	GM	0.04	78.76	0.03	0.17	20.09	0.33	0.58	98.85
	HQc	0.04	74.37	0.00	0.12	18.80	0.09	6.58	93.17
	HQ	0.04	73.39	0.00	0.09	18.49	0.06	7.94	91.88
	GoF	0.00	0.01	0.00	0.01	2.02	0.02	97.97	2.03
	R^2	0.00	20.18	0.00	0.01	1.66	0.02	78.16	21.83
	Adj R^2	0.03	55.39	0.00	0.04	13.62	0.01	30.91	69.01
	Q^2	0.70	35.39	3.78	0.73	34.71	5.67	19.23	70.10

Note: Success rate denotes the choice of Model 2 or 5.

Table 4B. Model Selection Rates Broken Down by Loading (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included

Loading condition	Criteria	Model #						Success rate
		1	2	3	4	6	7	
0.7	**AICc**	0.12	82.60	0.04	1.17	0.13	15.94	82.60
	AICu	0.09	88.13	0.07	1.20	0.42	10.10	88.13
	AIC	0.10	80.93	0.03	1.14	0.04	17.76	80.93
	CP	0.11	86.52	0.04	1.18	0.11	12.04	86.52
	FPE	0.10	80.96	0.03	1.14	0.04	17.73	80.96
	BIC	0.09	93.82	0.07	1.19	0.93	3.90	93.82
	GM	0.12	96.04	0.09	1.18	1.30	1.27	96.04
	HQ	0.09	89.08	0.07	1.19	0.30	9.29	89.08
	HQc	0.09	89.98	0.06	1.21	0.53	8.14	89.98
	GoF	0.10	0.30	0.07	0.02	0.04	99.54	0.30

Table 4B. Model Selection Rates Broken Down by Loading (Percentages)
Case 2: All Variables Included; Data-Generating Model Not Included

	R^2	0.02	35.60	0.00	0.64	0.06	63.72	35.60
	Adj R^2	0.07	68.42	0.01	0.96	0.03	30.53	68.42
	Q^2	7.11	32.63	11.86	5.33	30.66	12.50	32.63
0.8	**AICc**	0.04	84.21	0.02	0.64	0.11	14.97	84.21
	AICu	0.04	89.33	0.04	0.63	0.27	9.68	89.33
	AIC	0.04	82.50	0.02	0.60	0.06	16.78	82.50
	CP	0.04	88.50	0.03	0.63	0.10	10.69	88.50
	FPE	0.04	82.52	0.02	0.60	0.06	16.76	82.52
	BIC	0.04	95.14	0.04	0.61	0.54	3.61	95.14
	GM	0.04	97.17	0.08	0.63	0.82	1.26	97.17
	HQ	0.03	90.28	0.03	0.64	0.17	8.85	90.28
	HQc	0.04	91.41	0.04	0.63	0.29	7.58	91.41
	GoF	0.06	0.29	0.01	0.00	0.01	99.65	0.29
	R^2	0.02	32.48	0.00	0.31	0.00	67.20	32.48
	Adj R^2	0.03	69.72	0.00	0.49	0.01	29.75	69.72
	Q^2	3.20	45.42	9.41	2.48	23.70	15.91	45.42
0.9	**AICc**	0.04	85.48	0.00	0.10	0.03	14.35	85.48
	AICu	0.04	90.93	0.00	0.12	0.07	8.84	90.93
	AIC	0.04	83.67	0.00	0.07	0.01	16.22	83.67
	CP	0.04	91.90	0.00	0.09	0.02	7.95	91.90
	FPE	0.04	83.71	0.00	0.07	0.01	16.17	83.71
	BIC	0.04	96.92	0.00	0.14	0.13	2.76	96.92
	GM	0.04	98.83	0.03	0.17	0.33	0.59	98.83
	HQ	0.04	91.84	0.00	0.09	0.06	7.97	91.84
	HQc	0.04	93.14	0.00	0.12	0.09	6.60	93.14
	GoF	0.01	0.04	0.00	0.01	0.02	99.94	0.04
	R^2	0.00	21.19	0.00	0.01	0.02	78.80	21.19
	Adj R^2	0.03	68.77	0.00	0.04	0.01	31.15	68.77
	Q^2	0.71	68.97	3.78	0.73	5.67	20.34	68.97

Note: Success rate denotes the choice of Model 2.

Table 4C. Model Selection Rates Broken Down by Loading (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included

Loading condition	Criteria	Model #							Success rate
		1	2	3	4	5	6	7	
0.7	**AICc**	0.09	70.47	0.00	1.07	12.80	0.16	15.43	83.27
	AICu	0.11	74.19	0.03	1.06	13.84	0.36	10.41	88.03
	AIC	0.09	69.30	0.00	0.97	12.44	0.09	17.12	81.74

Table 4C. Model Selection Rates Broken Down by Loading (Percentages)
Case 3: Missing Variable (ξ_4); Data-Generating Model Not Included

	CP	0.09	73.33	0.00	1.04	13.57	0.11	11.86	86.90
	FPE	0.09	69.33	0.00	0.98	12.44	0.09	17.07	81.78
	BIC	0.11	79.28	0.06	0.99	14.91	0.86	3.80	94.19
	GM	0.13	80.92	0.10	1.08	15.24	1.20	1.32	96.17
	HQc	0.11	76.08	0.04	1.03	14.30	0.40	8.04	90.38
	HQ	0.09	75.14	0.01	1.08	14.16	0.24	9.28	89.30
	GoF	0.10	0.12	0.00	0.02	5.10	0.02	94.66	5.22
	R^2	0.02	34.53	0.00	0.62	2.41	0.03	62.40	36.94
	Adj R^2	0.08	59.59	0.00	0.83	10.09	0.03	29.39	69.68
	Q^2	6.28	17.69	11.34	5.19	17.00	31.76	10.97	34.69
0.8	**AICc**	0.07	69.24	0.04	0.74	14.70	0.07	15.14	83.94
	AICu	0.07	73.62	0.08	0.76	15.83	0.16	9.49	89.46
	AIC	0.06	67.83	0.03	0.68	14.36	0.07	16.98	82.19
	CP	0.08	72.83	0.04	0.78	15.69	0.08	10.50	88.52
	FPE	0.06	67.84	0.03	0.68	14.36	0.07	16.97	82.20
	BIC	0.06	77.71	0.08	0.78	17.13	0.53	3.71	94.84
	GM	0.04	79.25	0.11	0.80	17.60	0.91	1.28	96.85
	HQc	0.07	75.11	0.10	0.78	16.27	0.18	7.50	91.38
	HQ	0.07	74.28	0.07	0.76	15.97	0.10	8.77	90.24
	GoF	0.00	0.03	0.01	0.00	4.50	0.00	95.48	4.53
	R^2	0.01	31.13	0.01	0.41	2.50	0.00	65.95	33.63
	Adj R^2	0.04	58.27	0.01	0.58	11.61	0.01	29.48	69.88
	Q^2	3.06	24.03	9.42	3.03	22.59	23.60	14.46	46.62
0.9	**AICc**	0.01	68.02	0.04	0.10	17.06	0.02	14.74	85.08
	AICu	0.01	72.07	0.04	0.10	18.36	0.06	9.37	90.42
	AIC	0.01	66.64	0.02	0.09	16.62	0.02	16.59	83.27
	CP	0.01	72.78	0.04	0.12	18.52	0.01	8.51	91.30
	FPE	0.01	66.68	0.02	0.09	16.62	0.02	16.56	83.30
	BIC	0.02	77.18	0.04	0.13	19.46	0.11	3.06	96.63
	GM	0.02	79.12	0.04	0.16	19.90	0.26	0.50	99.02
	HQc	0.02	73.89	0.04	0.12	18.76	0.06	7.11	92.64
	HQ	0.01	72.67	0.04	0.10	18.49	0.02	8.67	91.16
	GoF	0.00	0.00	0.00	0.00	1.62	0.00	98.38	1.62
	R^2	0.00	19.13	0.00	0.01	1.73	0.00	79.12	20.87
	Adj R^2	0.01	55.46	0.01	0.08	13.58	0.00	30.87	69.03
	Q^2	0.57	35.17	3.66	0.57	34.33	5.99	19.91	69.50

Note: Success rate denotes the choice of Model 2 or 5.

6 Summary and Discussion

6.1 Key Findings and Recommendations

Our quest as IS researchers is to describe the sociotechnical processes that interest us with fidelity and economy. We create models to describe reality and uncover important relationships. Such models are a simplification or approximation of reality and hence cannot reflect it in its entirety (Burnham & Anderson, 2002). In fact, the only advantage a model may have over reality is its simplicity (Shugan, 2002). Models are important because they reflect some (partial) aspect of reality through a set of parameters and relationships by omitting distracting details. Such parameters have relevant, useful interpretations, even when they relate to quantities that are not directly observable (Burnham & Anderson, 2002). Importantly, there is not one unique model that characterizes the empirical evidence within a theoretical framework, but variations may offer theoretically justified alternatives for explaining the phenomenon under study. Comparing theoretically plausible models can therefore help address Grover's (2013, p. 5) concern that IS researchers have put "theories on a pedestal and treat them as immutable", and are quick to blame the method when hypotheses in a preferred model are not supported. From informal conversations with PLS researchers, we have learned that they in fact do explore several models, but eventually only report the results for a single model. The model selection approach thus supports this intuitive exploration and encourages researchers to report their alternative models rather than keep them "behind the scenes", leading to greater transparency. However, to reap the benefits of model comparisons, researchers must have adequate measures to identify the best model among a set of competing models. While corresponding criteria are routinely used in econometrics and factor-based structural equation modeling, the PLS framework did not have appropriate metrics for model selection to date.

Addressing this gap in research, this study introduces model selection criteria derived from information theory, to empirically assess alternative model configurations when using PLS. Specifically, we compared the efficacy of several model selection criteria across a range of model and data constellations that typically arise in practical situations. Our results clearly suggest that when comparing alternative models, PLS users can greatly benefit from using asymptotically consistent model selection criteria, which performed best in our simulation study. In particular, GM was the best-performing criterion with overall success rates of selecting a parsimonious yet well-fitting model in the upper 90s across all experimental conditions, closely followed by BIC. Thus, we recommend researchers to consider these criteria when performing theory-driven model comparisons. Researchers can easily compute these criteria manually using the formulae shown in Appendix B.[8] Furthermore, several model selection criteria analyzed in this study have been introduced in version 3.2.8 (and later) of the *SmartPLS 3* software (Ringle, Wende, & Becker, 2015),[9] and are also available in the R *semPLSic* package (Monecke, Sharma, & Kim, 2013), which complements the *semPLS* package (Monecke, 2012).

In contrast, the practice of model selection using R^2, Adjusted R^2, GoF, and Q^2 should be avoided. Not only did these criteria display a pronounced preference for the saturated model, but, in some cases (i.e., Q^2), they frequently selected incorrect models. That being said, the utility of PLS criteria lies in assessing a model's in-sample explanatory power (R^2) and predictive relevance (Q^2), which reflect other important aspects of model quality.

Because statistical inference assumes that a model has already been chosen, the roles and the sequence in which model comparison and model evaluation are conducted are different (Burnham & Anderson, 2002; Johnson & Omland, 2004). Thus, in line with Berk, Brown and Zhao (2010), we recommend the following five-step procedure for model selection and inference in a PLS study:

Step 1: A manageable set of theoretically plausible path models is developed.

Step 2: Data are collected.

Step 3: Measurement model assessment is performed for all models identified in Step 1.

Step 4: A model is selected based on the model selection criteria.

Step 5: Explanatory and predictive ability of the selected model is benchmarked, and statistical inference is applied to the structural paths to judge support for the proposed hypotheses.

After the researcher has clearly articulated and motivated the research issue, the first step is to identify and develop a manageable set of *theoretically justified* competing models that represent alternative explanations of the phenomenon (Burnham & Anderson, 2002). In general, there are multiple

[8] An Excel spreadsheet that illustrates the computation of all model selection criteria considered in this study using the standard output from any PLS software can be obtained from the Downloads section at https://www.pls-sem.net/.

[9] https://www.smartpls.com/documentation/algorithms-and-techniques/prediction-oriented-model-selection.

theoretical mechanisms that can explain the focal variable being studied (Imai & Tingley, 2012). The models selected for comparison should be motivated by theory from relevant fields, in line with PLS's "causal predictive" nature focused on theoretical explanation (Jöreskog and Wold, 1982, p. 270; Chin, 1995).[10] The researcher should leverage the existing literature to provide valid theoretical rationale for all the models being considered. In particular, the researcher should be able to (1) describe the theoretical commonalities among the proposed alternative models (i.e., whether certain proposed effects are common across models); (2) contrast the models to highlight the differences in theoretical mechanisms being captured (such differences may manifest in PLS as additional/different paths or antecedents); and (3) explain why the commonalities and differences are important to consider in terms of the effect on the target variable for the population under study.

The procedure continues with the data collection (Step 2), whose nature strongly depends on the aim of the research (Sarstedt, Bengrat, Shaltoni, & Lehmann, 2018). In this context, Calder, Phillips, and Tybout's (1981) distinction between *theory application* and *effects application* research is particularly relevant. The former type of research evaluates a specific theory about effects beyond a single research setting. Hence, "effects observed in the research are employed to assess the status of theory. But, it is the theoretical explanation that is expected to be generalizable and not the particular effects obtained" (Calder, Phillips, & Tybout 1981, p. 197). In this case, sample representativeness is of secondary concern when comparing models as long as specific sample characteristics are not an integral part of the theory (Calder et al., 1981). This differs from effects application, which "is based on a desire for knowledge about the events and relationships in a particular real-world situation. The primary goal of this type of research...is to obtain findings that can be applied directly to the situation of interest" (Calder et al., 1981, p. 198). Hence, when comparing alternative models in an effects application context, researchers need to ensure that their sample is representative of the specific population of interest.

Step 3 entails evaluating the reliability and validity of the measurement models by using criteria that have been well documented in textbooks (e.g., Ramayah et al., 2018; Hair, Hult, Ringle, & Sarstedt, 2017), tutorial articles (e.g., Chin, 2010; Gefen et al., 2011; Hair, Hollingsworth et al., 2017), and recent research articles (e.g., Cheah et al., 2018; Franke & Sarstedt, 2019; Henseler, Ringle, & Sarstedt, 2015). Researchers should use the same data treatment options (e.g., missing value treatments) and algorithm settings across the models to maintain consistency. This step ensures that only those models whose measures exhibit a sufficient degree of reliability and validity are considered further for model selection (Step 4).

Step 4 involves selecting a model based on the model selection criteria, followed by the structural model assessment and statistical inference for testing the research hypotheses (Step 5). While Steps 3 and 5 have been well documented in prior PLS research, our study focuses on model selection (Step 4), which had remained unaddressed to date. Hence, we issue some clarifying remarks to help researchers avoid falling prey to potential misconceptions regarding the use of model selection criteria.

First and foremost, model selection criteria are not meant to automate model selection with minimal thought. Rather, these criteria are meant as tools that researchers can use to gain additional information about models under consideration (Pitt & Myung, 2002). Selecting one model over another should proceed primarily on the basis of theoretical arguments aided by empirical evidence. Whether any of the models under consideration is theoretically meaningful depends primarily on the researcher's theoretical logic (Burnham & Anderson, 2002). Because PLS is used for theory building by focusing on explanation (Jöreskog & Wold, 1982; Chin, 1995), model selection using the thoughtless "all possible subsets" approach would be akin to "data dredging" (Burnham & Anderson, 2002, p. 37).

Second, like all empirical metrics, model selection criteria are also affected by the idiosyncrasies of the data at hand. The inability to select a "clear" winner is not a defect of the model selection criteria but rather an indication that the data are inadequate for such a strong inference (Burnham & Anderson, 2002). Science, as a discipline, rests on uncertainty both in theorizing and empirical modeling. The use of model selection criteria, while not a guarantee, brings transparency to the model selection process by providing researchers the ability to *acknowledge* model selection uncertainty.

Third, it is unrealistic to assume that there is one best criterion to use for all scenarios (Andrews & Currim, 2003b). For example, in Sarstedt et al.'s (2011) study on the performance of model selection criteria in the context of finite mixture-based latent class analysis in PLS, different degrees of model complexity had a significant effect on the criteria performance. However, while our recommendation to use BIC and GM certainly does not apply to *all* models, it is reasonable to conclude

[10] Because PLS path models focus on providing theoretical explanation, considering *purely* empirically motivated models would be akin to "snooping", and is not recommended for theoretical research that focuses on both explanation and prediction (Gregor, 2006).

that these metrics are useful for models with similar complexity as are commonly encountered in IS research and used in our simulation study.

Fourth, the model selection criteria can only be used to compare models on the same data set. Researchers cannot hope to compare and interpret models across samples using these criteria, as the properties of these data sets will be different. In particular, model selection criteria values cannot be compared across different missing data treatments commonly used in PLS (i.e., mean replacement, and casewise or pairwise deletion; Hair, Hult, Ringle, & Sarstedt, 2017). Similarly, all criteria assume that the same endogenous variable is the subject of interest; hence, it only makes sense to compare models with the same target construct (Burnham & Anderson, 2002).

Fifth, models selected by model selection criteria will often differ from those selected on the grounds of the null hypothesis testing paradigm (Posada & Buckley, 2004). In fact, the model selection criteria's performance is unrelated to the statistical significance of the path coefficients (Burnham & Anderson, 2002). While it may be sufficient (but not optimal) to rely on the null hypothesis paradigm in cases where only univariate causality is being determined, when multivariate causality is under consideration—as is the case in most PLS studies—statisticians agree that model comparison based on model selection criteria offers clear advantages (Elliott & Brook, 2007; Stephens, Buskirk, Hayward et al. 2005; Lukacs, Thompson, Kendall et al., 2007).

In Table 5, we summarize these and additional potential misconceptions and offer further clarifying remarks to avoid associated pitfalls as well as provide suitable references. Our illustration distinguishes between (1) research design issues, (2) differences with the null hypothesis paradigm, and (3) operational issues.

Table 5: Possible Misconceptions and Clarifications

Misconception	Clarification	Suggested readings
Research design issues		
Competing models are required in all PLS studies. A paper should be rejected if it does not compare models.	Competing models are not required in all PLS studies. Comparing alternative explanations (models) is a foundational step in the scientific method of strong inference whose application has encouraged faster innovation in more vigorous and livelier fields and separated them from lethargic ones. However, competing models have limited utility in strictly confirmatory settings where the goal is to judge the applicability of a *single* established model in a specific context.	Burnham & Anderson, 2001; Chamberlin, 1890; Clarke, 2007; Elliott & Brook, 2007; Nuzzo 2015; Platt, 1964; Yi & Nassen, 2015
Model selection criteria minimize (or eliminate) the role of theory and encourage comparing "atheoretical" or "thoughtless" models.	If all the models in the competing set are theoretically poor, the model selection criteria will still select a (poor) model (i.e., "garbage in, garbage out"). The effective use of model selection criteria requires researchers to generate a manageable set of *theoretically plausible* models by taking guidance from the existing theory.	Burnham & Anderson, 2001; Burnham & Anderson, 2002;
Model selection criteria will always lead to a single "best" model, and hence are a guarantee against model selection uncertainty. Researchers should always choose models that have the "best" selection criteria values.	Model selection criteria offer empirical guidance while comparing models. If a single model emerges as a clear winner, then inference can proceed based on that model. When there exist several models that are close in the information theoretic sense, and no clear winner emerges, then researchers should acknowledge model selection uncertainty. The tie among the models can then be broken to pick one model for *conditional* inference by considering the goal of the study, theoretical argumentation, and the full set of available empirical evidence (including explanatory and predictive metrics). In case *none* of the empirical criteria are able to separate the models, the models should be considered statistically equivalent. The choice of a model must then move beyond data analysis toward theoretical and practical considerations, as the data are unable to resolve the issue.	Buckland, Burnham & Augustin, 1997; Burnham & Anderson, 2002; Henley, Shook & Peterson, 2006; MacCallum, Wegener, Uchino et al. 1993; Raftery, 1995

Table 5: Possible Misconceptions and Clarifications

BIC and GM are the only recommendable model selection criteria across all scenarios.	While it is unrealistic to assume that there is one best criterion to use across all scenarios, BIC and GM are useful for models with similar complexity as commonly encountered in IS research and used in our simulation study.	Andrews & Currim, 2003b; Sarstedt, Becker, Ringle et al., 2011
Only parsimonious models should be selected. Alternatively, selecting a more complex model should always be avoided.	Parsimony is a key ingredient in *empirically* achieving a good balance between bias and variance in a specific data set and leads to generalizability and better out-of-sample predictive power. Parsimony has also been suggested as one of the primary goals of the scientific enterprise. The tradeoff between bias and variance is not related to theoretical complexity per se, which is expected to increase as more detailed knowledge about the system under consideration accumulates over time, and in this sense is a desirable trait.	Myung, 2000; Pitt & Myung, 2002; Simon, 2001; Shmueli & Koppius, 2011; Zellner, Keuzenkamp, & McAleer, 2001.
The true model is required to be in the competing set of models.	There is no mathematical requirement in the derivation of model selection criteria, including the asymptotically consistent criteria, which assumes the true model is in the competing set of models. In fact, a true model is generally out of reach and even worse, unknown. The results of this study show that the performance of the model selection criteria was unaffected by the inclusion or exclusion of the data-generating model in the competing set.	Burnham & Anderson, 2002; Burnham & Anderson, 2004; Cavanaugh & Neath, 1999
Comparison with the null hypothesis paradigm		
Model selection criteria are a replacement for *p*-values or measures of model fit. The use of *p*-values and PLS criteria is not recommended.	The method for selecting a model should be kept separate from the method used for model inference and reporting to avoid publication bias in PLS studies. The practice of *model selection* using path statistical significance (*p*-values) or PLS criteria is not recommended.	Berk et al., 2010; Cohen, 1994; Raftery, 1995
Models selected on the basis of path significance (*p*-values) or R^2 will be the same as those selected by model selection criteria. Therefore, the use of model selection criteria is redundant.	Because model selection criteria are tuned to the information content in a sample, their performance is unrelated to the statistical significance of path coefficients. In addition, PLS criteria such as R^2 try best to evaluate the model's fit to the "data", including its noise content. In contrast, model selection criteria try to model "information" contained in the data by avoiding noise and overfitting. This distinction means that models selected on the basis of model fit will tend to differ from models selected by model selection criteria, unless the data are totally devoid of noise. The overarching goal of science is not to simply look for statistical significance (or highest R^2) in a given sample, but rather meaningful relationships that best capture information in the data and generalize beyond the sample.	Burnham & Anderson, 2001; Burnham & Anderson, 2002; Johnson & Omland, 2004; Kirk, 1996; Pitt & Myung, 2002; Raftery, 1995; Reese, 2004
Model selection criteria will always select a model that is best in terms of the null hypothesis paradigm (or with greatest number of significant paths).	The use of path significance (*p*-values) offers no objective basis with which to judge which model is the best in the PLS framework and is not recommended for selecting a model among alternatives. In fact, the reliance on null hypothesis testing to select a model can induce publication bias via "p-value hacking" (see Section 2.2.1). The five-step procedure presented above delineates the roles of model selection and statistical inference to ensure that a model that offers the best balance between fit and parsimony in the given sample is selected *before* conducting statistical inference.	Aho, Derryberry, & Peterson, 2014; Raftery, 1995; Wasserstein & Lazar, 2016

Table 5: Possible Misconceptions and Clarifications

Model selection criteria can be used to state which model is "significantly" better than the other.	As model selection criteria are derived using information theory, researchers should avoid using statistical inference-related terminology associated with the null hypothesis testing paradigm (significant, type 1, type 2 errors, etc.) to avoid confusion. Model selection criteria do not provide evidence in the form of simple dichotomies such as "accept" or "reject", or "significant" or "nonsignificant". In particular, researchers cannot claim that one model is "significantly" better than another when using model selection criteria.	Burnham & Anderson, 2002; Johnson & Omland, 2004; Raftery, 1995
Operational issues		
Model selection criteria values can be compared across data sets. Or, the expectation that model selection criteria will always select the same model across data sets or contexts.	Model selection criteria can only be used to compare models on *exactly* the same data set. This also includes comparing the criteria across different missing data or outlier treatments commonly used in PLS as these methods change the underlying data. The onus is on the researcher to ensure that the sample is representative of the population under study. The generalizability of a model can then be assessed over time as more replication studies become available.	Burnham & Anderson, 2002;
Model selection criteria can be used to compare models with different target endogenous constructs.	All model selection criteria assume that the same target *endogenous* construct is the subject of interest. Hence, their use is limited to comparing models with the same target construct. However, model selection criteria do allow comparing models with different antecedent constructs.	Burnham & Anderson, 2002
Model selection criteria can be used with single models. Alternatively, standalone values of model selection criteria are interpretable and useful.	Because model selection criteria are *relative* distance metrics, a standalone value, no matter how large or small, is not meaningful. Thus, interpreting a model selection criterion value for a single model is meaningless. Instead, the difference in values of two models can be interpreted as the relative loss in information between them.	Burnham & Anderson, 2001; Burnham & Anderson, 2002; Burnham & Anderson, 2004; Raftery, 1995
Model selection criteria have values restricted to a specific range.	Unlike R^2, which varies between 0 and 1 and has a useful interpretation, the model selection criteria do not have a scale. Thus, a wide range of values (including negative values) are possible. Furthermore, there are no strict "cut-off" values to indicate which models are important.	Burnham & Anderson, 2002; Johnson & Omland, 2004; Raftery, 1995
Model selection criteria can only be used to compare nested models.	In contrast to the null hypothesis testing paradigm, model selection criteria are not bound by the nesting requirement and can be used to select among nonnested models.	Burnham & Anderson, 2002; Johnson & Omland, 2004; Raftery, 1995
The order in which models are compared matters.	The order of comparison when using model selection criteria is irrelevant, unlike the hypothesis testing approaches (e.g., forward addition, backward elimination, and stepwise) that may be affected by the order.	Burnham & Anderson, 2002; Johnson & Omland, 2004; Raftery, 1995

6.2 Limitations and Future Research

As with all research, our study is subject to limitations that offer avenues for future research. First, because PLS estimation does not necessarily maximize the overall likelihood function, we have adapted and evaluated the use of model selection criteria in this context under a variety of scenarios. While our results show consistency across different manipulations, we have investigated only a subset of all possible model selection scenarios, and it is not guaranteed that the same behavior will extend to other cases. Specifically, exploring the performance of these model selection criteria in more complex models involving interaction effects (e.g., Henseler & Chin, 2010), nonlinear effects (e.g., Rigdon, Ringle, & Sarstedt, 2010), or hierarchical component models (e.g., Becker, Klein, & Wetzels, 2012) would be particularly fruitful. In doing so, researchers should examine a broader set of data constellations, by considering, for example, lateral collinearity (Kock & Lynn, 2012).

Second, while beyond the scope of this study, model selection criteria also allow researchers to *quantify* model selection uncertainty through the use of model ranks and Akaike weights (Dayton, 2003; Johnson & Omland, 2004). Akaike weights quantify the evidence in favor of a model being the actual best model in comparison to other models for the situation at hand (Burnham & Anderson, 2002). By comparing the rank and Akaike weights of all competing models, researchers can create evidence ratios and judge the relative strength of evidence for each participating model in comparison to others (Wagenmakers & Farrell, 2004). More importantly, Akaike weights allow researchers to create more accurate model-averaged estimates based on a set of models (i.e., multimodel inference) rather than relying on a single model, which can improve out-of-sample predictive accuracy (Burnham & Anderson, 2004; Claeskens & Hjort, 2008; Hansen, 2007; Symonds & Mousalli, 2011). We believe that this can be a particularly fruitful area for future research to explore.

Third, because model selection and PLS criteria require focus on a specific endogenous construct, comparing two theories with different target constructs is not feasible. However, Cohen, Carlsson, Ballesteros, and Amant (1993) have proposed an extension to path analysis that allows for comparing models with reversed causality and which has been successfully implemented in a PLS context (Sattler, Völckner, Riediger, & Ringle, 2010). Future research should try to merge these two approaches to broaden the focus of model selection.

Finally, we have restricted our attention to reflectively specified measures and further research would be necessary to examine the case of formatively specified constructs, which have received considerable attention in the recent IS research (e.g., Diamantopoulos 2011; Petter, Straub, & Rai, 2007).

7 Conclusion

The overall goal of this paper is to encourage model selection practices in PLS with the aim of helping create generalizable theories in IS research. The practice of model selection has been considered a fundamental building block of scientific progress by the philosophers of science (e.g., Lakatos, 1970; Meehl, 1990), and is consistent with the fundamental scientific pursuit of strong inference (Platt, 1964). While models, by definition, cannot reflect all of reality (Burnham & Anderson, 2002), they can be used to *approximate* the mechanisms that underlie the data generation process (MacCallum, 2003). In practice, this means that researchers should seek to identify a model that is parsimonious and consistent with reality. Such a model achieves a sound balance between fit to the data and parsimony and, hence, has the best chance of being generalizable. Theory guides the derivation of models that are likely to reflect reality adequately but evaluating them in isolation is likely to be prone to confirmation bias. Having a set of theory-driven competing models is one way to circumvent this bias (Preacher, 2006).

The ability to evaluate theory-driven (plausible) competing models is especially relevant in exploratory settings, a context for which PLS was originally designed (e.g., Wold, 1974; 1980). Our study compared the performance of model selection criteria in PLS under a variety of data and model constellations. The results strongly suggest that researchers should avoid basing the decision to select a model using the PLS criteria (i.e., R^2, Adjusted R^2, GoF, and Q^2), as is the current practice in academic research and industry. Instead, the use of model selection criteria (in particular, BIC and GM) is advised due to their robust model selection performance under all the contexts studied here. With a proper theoretical base and a strong study design, these criteria allow researchers to consider competing models within the PLS framework and select the best model among them. While certainly not a panacea, we believe that these criteria can provide the much-needed empirical guidance to researchers to make more informed model selection decisions under exploratory settings.

Acknowledgments

The authors are grateful for the guidance provided by Dr. Stacie Petter and the anonymous reviewers. The third author was partially supported by the Ministry of Science and Technology, Taiwan (Grant 105-2410-H-007-034-MY3). This paper is dedicated to the memory of the late Dr. Kevin H. Kim who helped lay the foundation for this work.

References

Addas, S. (2010). A call for engaging context in HCI/MIS research with examples from the area of technology interruptions. *AIS Transactions on Human-Computer Interaction*, *2*(4), 178-196.

Aguirre-Urreta, M. I., & Marakas, G. M. (2012). Revisiting bias due to construct misspecification: Different results from considering coefficients in standardizes form. *MIS Quarterly*, *36*(1), 123-138.

Aguirre-Urreta, M., & Rönkkö, M. (2018). Statistical inference with PLSc using bootstrap confidence intervals. *MIS Quarterly*, *42*(3), 1001-1020.

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, *22*(1), 203-217.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199-213). New York, NY: Springer.

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631-636.

Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M. & Ryu, K. (2018), An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*, *30*(1), 514-538.

Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review*, *32(4)*, 1265-1281.

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Managemen*t, *64*(4), 912-923.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411-423.

Anderson, J. C., Gerbing, D. W., & Hunter, J. E. (1987). On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria. *Journal of Marketing Research*, *24*(4), 432-437.

Andrews, R. L., & Currim, I. S. (2003a). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing*, *20*(4), 315-321.

Andrews, R. L., & Currim, I. S. (2003b). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, *40*(2), 235-243.

Angst, C. M., Devaraj, S., & D'Arcy, J. (2012). Dual role of IT-assisted communication in patient care: a validated structure-process-outcome framework. *Journal of Management Information Systems*, *29*(2), 257-292.

Arazy, O., & Gellatly, I. R. (2012). Corporate wikis: The effects of owners' motivation and behavior on group members' engagement. *Journal of Management Information Systems*, *29*(3), 87-116.

Armstrong, D. J., Brooks, N. G., & Riemenschneider, C. K. (2015). Exhaustion from information system career experience: Implications for turn-away intention. *MIS Quarterly*, *39*(3), 713-727.

Becker, J. M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning*, *45*(5), 359-394.

Becker, J.-M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats. *MIS Quarterly*, *37*(3), 665-694.

Becker, J. M., Ringle, C. M., Sarstedt, M., & Völckner, F. (2015). How collinearity affects mixture regression results. *Marketing Letters*, *26*(4), 643-659.

Benlian, A., Koufaris, M., & Hess, T. (2011). Service quality in software-as-a-service: Developing the SaaS-Qual measure and examining its role in usage continuance. *Journal of Management Information Systems*, *28*(3), 85-126.

Bentler, P. M., & Mooijaart, A. B. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin*, *106*(2), 315-317.

Berk, R. A. (2008). *Statistical learning from a regression perspective*. New York, NY: Springer.

Berk, R., Brown, L., & Zhao, L. (2010). Statistical inference after model selection. *Journal of Quantitative Criminology*, *26(2)*, 217-236.

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, *22*(7), 719-725.

Bozdogan, H. (1994). Mixture-model cluster analysis using model selection criteria and a new informational measure of complexity. In H. Bozdogan (Ed.), *Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach* (Vol. 2, pp. 69-113). Boston, MA: Kluwer Academic Publishers.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, 53*(2)*, 603-618.

Burnham, K. P., & Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, *28(2)*, 111-119.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261-304.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23-35.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, *8*(2), 197-207.

Cavanaugh, J. E., & Neath, A. A. (1999). Generalizing the derivation of the Schwarz information criterion. *Communications in Statistics-Theory and Methods*, *28*(1), 49-66.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195-212.

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, *15*(366), 92-96.

Chandra, S., Srivastava, S. C., & Theng, Y. L. (2012). Cognitive absorption and trust for workplace collaboration in virtual worlds: An information processing decision making perspective. *Journal of the Association for Information Systems*, *13*(10), 797-835.

Cheah, J.-H., Sarstedt, M., Ringle, C. M., Ramayah, T., & Ting, H. (2018). Convergent validity assessment of formatively measured constructs in PLS-SEM: On using single-item versus multi-item measures in redundancy analyses. *International Journal of Contemporary Hospitality Management*, *30*(11), 3192-3210.

Chin, W. W. (1995). Partial least squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, *2*(2), 315-319.

Chin, W. W. (2010). How to write up and report PLS analyses. In V. Esposito Vinzi, W. W. Chin, J. Henseler & H. Wang, et al. (Eds.), *Handbook of partial least squares. Concepts, methods and application*s (pp. 655-690). Berlin: Springer.

Chin, W. W., Thatcher, J. B., & Wright, R. T. (2012). Assessing common method bias: Problems with the ULMC technique. *MIS Quarterly*, *36*(3), 1003-1019.

Chin, W. W., & Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: A note of caution. *MIS Quarterly*, *19*(2), 237-246.

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge, UK: Cambridge University Press.

Clarke, K. A. (2007). The necessity of being comparative: Theory confirmation in quantitative political science. *Comparative Political Studies*, *40(7)*, 886-908.

Cohen, J. (1994). The earth is round (p< .05). *American Psychologist*, *49*(12), 997-1003.

Cohen, P. R., Carlsson, A., Ballesteros, L., & Amant, R. S. (1993). Automating path analysis for building causal models from data. *Proceedings of the International Workshop on Machine Learning* (pp. 57-64).

D'Arcy, J., Herath, T., & Shoss, M. K. (2014). Understanding employee responses to stressful information security requirements: a coping perspective. *Journal of Management Information Systems*, *31*(2), 285-318.

Dayton, C. M. (2003). Model comparisons using information measures. *Journal of Modern Applied Statistical Methods*, *2*(2), 281-292.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335-358.

Dijkstra, T. & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297-316.

Dill, K. E. (2013). *The Oxford handbook of media psychology*. Oxford, UK: Oxford University Press.

Dinger, M., Thatcher, J. B., Treadway, D., Stepina, L., & Breland, J. (2015). Does professionalism matter in the IT workforce? An empirical examination of IT professionals. *Journal of the Association for Information Systems*, *16*(4), 281-313.

Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*(8746), 867-872.

Elie-Dit-Cosaque, C., Pallud, J., & Kalika, M. (2011). The influence of individual, contextual, and social factors on perceived behavioral control of information technology: A field theory approach. *Journal of Management Information Systems*, *28*(3), 201-234.

Elliott, L. P., & Brook, B. W. (2007). Revisiting Chamberlin: Multiple working hypotheses for the 21st century. *BioScience*, 57(7), 608-614.

Evermann, J. & Tate, M. (2010). Testing models or fitting models? Identifying model misspecifications in PLS. *Proceedings of the 31st International Conference on Information Systems*.

Fang, Y., Qureshi, I., Sun, H., McCole, P., Ramsey, E., & Lim, K. H. (2014). Trust, satisfaction, and online repurchase intention: The moderating role of perceived effectiveness of e-commerce institutional mechanisms. *MIS Quarterly*, *38*(2), 407-427.

Forster, M. R. (2000). Hard problems in the philosophy of science: Idealization and commensurability. In R. Nola & H. Sankey (Eds.), *After Popper, Kuhn and Feyerabend: Issues in theories of scientific method* (pp. 231-250). Berlin: Springer.

Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: A comparison of four procedures. *Internet Research*, forthcoming.

Friedman, M. (1994). The methodology of positive economics. *The Philosophy of Economics: An Anthology*, *2*, 180-213.

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

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*(2), 186-192.

Geweke, J., & Meese, R. (1981). Estimating regression models of finite but unknown order. *International Economic Review*, *22*(1), 55-70.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly, 36*(3), 981-1001.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611-642.

Grgecic, D., Holten, R., & Rosenkranz, C. (2015). The Impact of functional affordances and symbolic expressions on the formation of beliefs. *Journal of the Association for Information Systems*, *16*(7), 580-607.

Grover, V. (2013). Muddling along to moving beyond in IS research: Getting from good to great. *Journal of the Association for Information Systems*, *14*(5), 274-282.

Grover, V., Lyytinen, K., Srinivasan, A., & Tan, B. C. (2008). Contributing to rigorous and forward thinking explanatory theory. *Journal of the Association for Information Systems*, *9(2)*, 40-47.

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing*, forthcoming.

Hair J. F., 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*, *117*(3), 442-458.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Thousand Oaks, CA: SAGE.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science*, *45*(5), 616-632.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. J*ournal of the Academy of Marketing Science*, *40(3)*, 414-433.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*(2), 190-195.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, *75*(4), 1175-1189.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference. and Prediction* (2nd ed.). New York, NY: Springer.

Hawkins, D. S., Allen, D. M., & Stromberg, A. J. (2001). Determining the number of components in mixtures of linear models. *Computational Statistics & Data Analysis*, *38*(1), 15-48.

Henley, A. B., Shook, C. L., & Peterson, M. (2006). The presence of equivalent models in strategic management research using structural equation modeling: Assessing and addressing the problem. *Organizational Research Methods*, *9(4)*, 516-535.

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., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, *28*(2), 565-580.

Henseler, J, Ringle, C. M., & Sarstedt, M. (2015). New criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135.

Homburg, C. (1991). Cross-validation and information criteria in causal modeling. *Journal of Marketing Research*, *28*(2), 137-144.

Hsieh, J. P. A., Rai, A., Petter, S., & Zhang, T. (2012). Impact of user satisfaction with mandated CRM use on employee service quality. *MIS Quarterly*, *36*(4), 1065-1080.

Imai, K., & Tingley, D. (2012). A statistical method for empirical testing of competing theories. *American Journal of Political Science*, *56(1)*, 218-236.

Iyengar, K., Sweeney, J. R., & Montealegre (2015). Information technology use as a learning mechanism: the impact of IT use on knowledge transfer effectiveness, absorptive capacity, and franchisee performance. *MIS Quarterly*, *39*(3), 615-641.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of management review*, *31(2)*, 386-408.

Johnson, N. A., & Cooper, R. B. (2015). Understanding the influence of instant messaging on ending concessions during negotiations. *Journal of Management Information Systems*, *31*(4), 311-342.

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101-108.

Johnston, A. C., Warkentin, M., & Siponen, M. T. (2015). An enhanced fear appeal rhetorical framework: leveraging threats to the human asset through sanctioning rhetoric. *MIS Quarterly*, *39*(1), 113-134.

Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects, in H. Wold and K. Jöreskog (Eds.), *Systems under indirect observation: causality, structure, prediction* (Vol. 1, pp. 263-270), Amsterdam: Elsevier.

Kankanhalli, A., Ye, H., & Teo, H. H. (2015). Comparing potential and actual innovators: an empirical study of mobile data services innovation. *MIS Quarterly*, *39*(3), 667-682.

Karahanna, E., & Preston, D. S. (2013). The effect of social capital of the relationship between the CIO and top management team on firm performance. *Journal of Management Information Systems*, *30*(1), 15-56.

Khan, G. F., Sarstedt, M., Shiau, W.-L., Hair, J. F., Ringle, C. M., & Fritze, M. P. (2019). Methodological research on partial least squares structural equation modeling (PLS-SEM). An analysis based on social network approaches. *Internet Research*, forthcoming.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746-759.

Knaus, J. (2013). Snowfall: Easier cluster computing (based on snow). R package version 1.84-6. Retrieved from https://cran.r-project.org/web/packages/snowfall/index.html

Kock, N., & Chatelain-Jardón, R. (2011). Four guiding principles for research on evolved information processing traits and technology-mediated task performance. *Journal of the Association for Information Systems*, *12*(10), 684-713.

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

Konishi, S., & Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, *83*(4), 875-890.

Kumar, A., & Sharma, S. (1999). A metric measure for direct comparison of competing models in covariance structure analysis. *Structural Equation Modeling*, *6*(2), 169-197.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In S. G. Harding (Ed.), *Can theories be refuted? Essays on the Duhem-Quine thesis* (pp. 205-259). Berlin: Springer.

Lauenroth, W. K. (2003). *Models in ecosystem science*. Princeton, NJ: Princeton University Press.

Lee, L., Petter, S., Fayard, D., & Robinson, S. (2011). On the use of partial least squares path modeling in accounting research. *International Journal of Accounting Information Systems*, *12*(4), 305-328.

Li, B., Morris, J., & Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *64*(1), 79-89.

Lin, M., Lucas Jr., H. C., & Shmueli, G. (2013). Research commentary-too big to fail: Large samples and the p-value problem. *Information Systems Research*, *24*(4), 906-917.

Lin, L.C., Huang, P.-H., & Wenig, L.J. (2017). Selecting path models in SEM: A comparison of model selection criteria. *Structural Equation Modeling*, *24*(6), 855-869.

Lukacs, P. M., Thompson, W. L., Kendall, W. L., Gould, W. R., Doherty, P. F., Burnham, K. P., & Anderson, D. R. (2007). Concerns regarding a call for pluralism of information theory and hypothesis testing. *Journal of Applied Ecology*, *44*(2), 456-460.

MacCallum, R. C. (2003). 2001 presidential address: Working with imperfect models. *Multivariate Behavioral Research*, *38*(1), 113-139.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114(1)*, 185.

Majchrzak, A., Wagner, C., & Yates, D. (2013). The impact of shaping on knowledge reuse for organizational improvement with wikis. *MIS Quarterly*, *37*(2), 455-469.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, *15*(4), 661-675.

Matook, S., Cummings, J., & Bala, H. (2015). Are you feeling lonely? The impact of relationship characteristics and online social network features on loneliness. *Journal of Management Information Systems*, *31*(4), 278-310.

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

McQuarrie, A. D., & Tsai, C. L. (1998). *Regression and time series model selection* (vol. 43). Singapore: World Scientific.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108-141.

Mehta, N., & Bharadwaj, A. (2015). Knowledge integration in outsourced software development: the role of sentry and guard processes. *Journal of Management Information Systems*, *32*(1), 82-115.

Monecke, A. (2012). semPLS: An R package for structural equation models using partial least squares. R Package Version 1.0-08. Retrieved from https://cran.r-project.org/web/packages/semPLS/index.html

Monecke, A., Sharma, P. N., & Kim, K. H. (2013). semPLSic. Information Criteria for PLS based SEM. R Package Version 0.1-0. Retrieved from https://r-forge.r-project.org/scm/viewvc.php/pkg/semPLSic/DESCRIPTION?root=sempls&view=log

Moon, J. W., & Kim, Y. G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217-230.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*(1), 190-204.

Myung, J. I., & Pitt, M. A. (2004). Model comparison methods. *Methods in Enzymology*, *383*, 351-366.

Navarro, D. J., & Myung, J. I. (2005). Model evaluation. *Encyclopedia of Statistics in Behavioral Science*.

Nuzzo, R. (2015). How scientists fool themselves—and how they can stop. *Science*, *526*(7572), 182-185.

Ou, C. X., Pavlou, P. A., & Davison, R. (2014). Swift guanxi in online marketplaces: the role of computer-mediated communication technologies. *MIS Quarterly*, *38*(1), 209-230.

Pan, W. (2001a). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57(1), 120-125.

Pan, W. (2001b). Model selection in estimating equations. *Biometrics*, *57*(2), 529-534.

Park, I, Sharman, R., & Rao H. R. (2015). Disaster experience and hospital information systems: An examination of perceived information assurance, risk, resilience, and HIS usefulness. *MIS Quarterly*, *39*(2), 317-344.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*(2), 287-312.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31(4)*, 623-656.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421-425.

Platt, J. R. (1964). Strong inference. *Science*, *146* (3642), 347-353.

Plouffe, C. R., Hulland, J. S., & Vandenbosch, M. (2001). Research report: Richness versus parsimony in modeling technology adoption decisions—understanding merchant adoption of a smart card-based payment system. *Information Systems Research*, *12*(2), 208-222.

Polites, G. L., & Karahanna, E. (2012). Shackled to the status quo: The inhibiting effects of incumbent system habit, switching costs, and inertia on new system acceptance, *MIS Quarterly*, *36*(1), 21-41.

Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, *23*(10), 1891-1901.

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793-808.

Preacher, K. J. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research*, *41*(3), 227-259.

Popper, K. R. (1959). *The logic of scientific discovery*. *London: Hutchinson*.

Purcell, W. M. (1992). Are there so few communication theories? *Communication Monographs*, *59*(1), 94-97.

R Development Core Team. (2014). R: A language and environment for statistical computing. The R foundation for statistical computing, Vienna, Austria. Available online at https://www.R-project.org/

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163.

Ramayah, T., Cheah, J., Chuah, H.T. and Memon M.A. (2018). *Partial least squares structural equation modeling (PLS-SEM) using SmartPLS 3.0*, (2nd ed.). Kuala Lumpur: Pearson Malaysia.

Ray, S., Kim, S. S., & Morris, J. G. (2012). Research note: Online users' switching costs: Their nature and formation. *Information Systems Research*, *23*(1), 197-213.

Raymond, L. (1985). Organizational characteristics and MIS success in the context of small business. *MIS Quarterly, 9*(1), 37-52.

Reese, R. A. (2004). Does significance matter? *Significance*, *1*(1), 39-40.

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

Rigdon, E. E., Ringle, C. M., & Sarstedt, M. (2010). Structural modeling of heterogeneous data with partial least squares. *Review of Marketing Research*, *7*, 255-296.

Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing ZFP – Journal of Research and Management*, *39*(3), 4-16.

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

Ringle, C. M., Sarstedt, M., & Schlittgen, R. (2014). Genetic algorithm segmentation in partial least squares structural equation modeling. *OR Spectrum*, *36*(1), 251-276.

Ringle, C., Wende, S., & Becker, J. M. (2015). SmartPLS 3.0, Hamburg. Available at http://www.smartpls.de.

Ringle, C. M., Sarstedt, M., Mitchell, R. & Gudergan, S. P. (2019). Partial least squares structural equation modeling in HRM research. *International Journal of Human Resource Management*, forthcoming.

Roberts, N., & Grover, V. (2009). Theory development in information systems research using structural equation modeling: Evaluation and recommendations. In Y. K. Dwivedi, B. Lal, M. D. Williams, S. L. Schneberger, & M. Wade (Eds.), *Handbook of research on contemporary theoretical models in information systems* (pp. 77-94). Hershey, PA: IGI Global.

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

Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences*, *87*, 76-84.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641.

Rust, R. T., Simester, D., Brodie, R. J., & Nilikant, V. (1995). Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Management Science*, *41*(2), 322-333.

Rust, R. T., Lee, C., & Valente Jr., E. (1999). Comparing covariance structure models: A general methodology. *International Journal of Research in Marketing*, *12*(4), 279-291.

Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments? *Schmalenbach Business Review*, *63*(1), 34-62.

Sarstedt, M., Wilczynski, P., & Melewar, T. C. (2013). Measuring reputation in global markets: A comparison of reputation measures' convergent and criterion validities. *Journal of World Business*, *48*(3), 329-339.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998-4010.

Sarstedt, M., Ringle, C. M., & Hair, J.F. (2017). Partial least squares structural equation modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.). *Handbook of Market Research*. Heidelberg, Germany: Springer.

Sarstedt, M., Bengart, P., Shaltoni, A. M., & Lehmann, S. (2018). The use of sampling methods in advertising research: A gap between theory and practice. *International Journal of Advertising*, *37*(4), 650-663.

Sattler, H., Völckner, F., Riedigerm, C., & Ringle, C. M. (2010). The impact of brand extension success drivers on brand extension price premiums. *International Journal of Research in Marketing*, *27*(4), 319-328.

Schlittgen, R. (2015). SEGIRLS: Clusters regression, pls path and gsca models by iterative reweighting. R package version 0.5. Available at http://www3.wiso.uni-hamburg.de/fileadmin/bwl/statistikundoekonometrie/Schlittgen/SoftwareUndDaten/SEGIRLS_0.5.tar.gz

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Setia, P., Venkatesh, V., & Joglekar, S. (2013). Leveraging digital technologies: how information quality leads to localized capabilities and customer service performance. *MIS Quarterly*, *37*(2), 565-590.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, *8*(1), 147-164.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289-310.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553-572.

Shugan, S. M. (2002). Editorial: Marketing science, models, monopoly models, and why we need them. *Marketing Science*, *21*(3), 223-228.

Simon, H. A. (2001). Science seeks parsimony, not simplicity: searching for pattern in phenomena. In A. Zellner, H. A. Keuzenkamp, & M. McAleer (Eds.). *Simplicity, inference and modelling: Keeping it sophisticatedly simple* (pp. 32-72). Cambridge, UK: Cambridge University Press.

Sin, C. Y., & White, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, *71*(1), 207-225.

Sobel, M. (2000). Causal inference in the social sciences. *Journal of American Statistical Association*, *95(450)*, 647-651.

Spiess, A.-N., & Neumeyer, N. (2010). An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. *BMC Pharmacology*, *10*, 6.

Stephens, P. A., Buskirk, S. W., Hayward, G. D., & Martinez Del Rio, C. (2005). Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology*, *42*(1), 4-12.

Sugiura, N. (1978). Further analysts of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, *7*(1), 13-26.

Sun, H. (2012). Understanding user revisions when using information system features: Adaptive system use and triggers. *MIS Quarterly*, *36*(2), 453-478.

Sykes, T. A. (2015). Support structures and their impacts on employee outcomes: A longitudinal field study of an enterprise system implementation. *MIS Quarterly*, *39*(2), 473-495.

Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13-21.

Tan, C. W., Benbasat, I., & Cenfetelli, R. T. (2013). IT-mediated customer service content and delivery in electronic governments: An empirical investigation of the antecedents of service quality. *MIS Quarterly*, *37*(1), 77-109.

Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). A global goodness-of-fit index for PLS structural equation modelling. *Proceedings of the XLII SIS scientific meeting* (Vol. 1, pp. 739-742).

Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*(2), 76-92.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478.

Venkatesh, V., Brown, S. A., Maruping, L. M., & Bala, H. (2008). Predicting different conceptualizations of system use: The competing roles of behavioral intention, facilitating conditions, and behavioral expectation. *MIS Quarterly*, *32*(3), 483-502.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157-178.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228-243.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779-804.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192-196.

Wagner, H. T., Beimborn, D., & Weitzel, T. (2014). How social capital among information technology and business units drives operational alignment and IT business value. *Journal of Management Information Systems*, *31*(1), 241-272.

Wan, Z., Compeau, D., & Haggerty, N. (2012). The effects of self-regulated learning processes on e-learning outcomes in organizational settings. *Journal of Management Information Systems*, *29*(1), 307-340.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129-133.

Wit, E., Heuvel, E. V. D., & Romeijn, J. W. (2012). "All models are wrong...": An introduction to model uncertainty. *Statistica Neerlandica*, *66* (3), 217-236.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16* (1), 85-102.

Wold, H. (1974). Causal flows with latent variables: Partings of the ways in the light of NIPALS modelling. *European Economic Review*, *5*(1), 67-86.

Wold, H. (1980). Model construction and evaluation when theoretical knowledge is scarce. In J. Kmenta & J. B. Ramsey (Eds.), *Evaluation of econometric models* (pp. 47-74). New York, NY: Academic.

Wold, H. (1982). Soft modeling: the basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1-54). Amsterdam: Elsevier

Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Volume 6, pp. 581-591). New York, NY: Wiley.

Xu, J. D., Benbasat, I., & Cenfetelli, R. (2011). The effects of service and consumer product knowledge on online customer loyalty. *Journal* *of the Association for Information Systems*, *12*(11), 741-766.

Xu, S. X., & Zhang, X. (2013). Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly*, *37*(4), 1043-1068.

Xue, L., Zhang, C., Ling, H., & Zhao, X. (2013). Risk mitigation in supply chain digitization: System modularity and information technology governance. *Journal of Management Information Systems*, *30*(1), 325-352.

Xue, Y., Liang, H., & Wu, L. (2011). Punishment, justice, and compliance in mandatory IT settings. *Information Systems Research*, *22*(2), 400-414.

Yafune, A., Funatogawa, T., & Ishiguro, M. (2005). Extended information criterion (EIC) approach for linear mixed effects models under restricted maximum likelihood (REML) estimation. *Statistics in Medicine*, *24*(22), 3417-3429.

Yamaoka, K., Nakagawa, T., & Uno, T. (1978). Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *Journal of Pharmacokinetics and Pharmacodynamics*, *6*(2), 165-175.

Yi, Y., & Nassen, K. (2015). Multiple comparison and cross-validation in evaluating structural equation models. In V. L. Crittenden (Ed.), *Proceedings of the 1992 Academy of Marketing Science (AMS) Annual Conference* (pp. 407-411). Berlin: Springer.

Zhang, X., Venkatesh, V., & Brown, S. A. (2011). Designing collaborative systems to enhance team performance. *Journal of the Association for Information Systems*, *12*(8), 556-584.

Zellner, A., Keuzenkamp, H. A., & McAleer, M. (Eds.). (2001). *Simplicity, inference and modelling: Keeping it sophisticatedly simple*. Cambridge, UK: Cambridge University Press.

Appendix A

Table A1. PLS-Based Studies that Analyzed Multiple Models for Reasons Other than Comparing Theoretically Justified Alternatives in the Top Four IS Journals (*ISR*, *JAIS*, *JMIS*, and *MISQ*) Published April 2011 - April 2016.

Reference	Aim of the analysis involving multiple models
Grgecic, Holten, & Rosenkranz (2015) | *JAIS*	Alternative construct operationalizations
Karahanna & Preston (2013) | *JMIS*	Alternative construct operationalizations
Ray, Kim, & Morris (2012) | *ISR*	Alternative construct operationalizations
Elie-Dit-Cosaque, Pallud, & Kalika (2011) | *JMIS*	Comparison with another research study
Xu, Benbasat, & Cenfetelli (2013) | *MISQ*	Explanatory power when adding antecedent construct
Fang et al. (2014) | *MISQ*	Moderator model
Ou, Pavlou, & Davison (2014) | *MISQ*	Moderator model
Sun (2012) | *MISQ*	Moderator model
Venkatesh, Thong, & Xu (2012) | *MISQ*	Moderator model
Xu, Benbasat, & Cenfetelli (2011) | *JAIS*	Moderator model
Matook, Cummings, & Bala (2015) | *JMIS*	Only controls, main effects, main effects plus interactions (stability check)
Mehta & Bharadwaj (2015) | *JMIS*	Only controls, main effects, main effects plus interactions (stability check)
Angst, Devaray, & D'Arcy (2012) | *JMIS*	Stability when antecedent constructs removed
D'Arcy, Herath, & Shoss (2014) | *JMIS*	Stability when antecedent constructs removed
Johnson & Coopper (2015) | *JMIS*	Stability when antecedent constructs removed
Xue. Liang, & Wu (2011) | *ISR*	Stability when antecedent constructs removed
Armstrong, Brooks & Riemenschneider (2015) | *MISQ*	Stability when comparing with saturated model
Arazy & Gellatly (2012) | *JMIS*	Stability when control variables included
Kock & Chatelain-Jardón (2011) | *JAIS*	Stability when control variables included
Setia, Venkatesh, & Joglekar (2013) | *MISQ*	Stability when control variables included
Zhang, Venkatesh, & Brown (2011) | *JAIS*	Stability when control variables included
Wan, Compeau, & Haggerty (2012) | *JMIS*	Stability when control variables removed
Kankanhalli, Ye, & Teo (2015) | *MISQ*	Stability when different endogenous construct
Benlian, Koufaris, & Hess (2011) | *JMIS*	Stability when structural paths added
Johnston, Warkentin, & Siponen (2015) | *MISQ*	Stepwise extension of model to test stability
Note: JAIS=Journal of the Association for Information Systems, JMIS=Journal of Management Information Systems, MISQ=MIS Quarterly	

Table A2. PLS-Based Studies that Compared Theoretically Justified Alternative Models in the Top Four IS Journals (*ISR*, *JAIS*, *JMIS*, and *MISQ*) Published April 2011 - April 2016.

Reference	Criterion used for model selection
Chandra, Srivastava, & Theng (2012) | *JAIS*	f^2, R^2
Dinger et al. (2015) | *JAIS*	R^2
Hsieh, Rai, Petter, & Zhang (2012) | *MISQ*	R^2
Majchrzak, Wagner, & Yates (2013) | *MISQ*	R^2
Polites & Karahanna (2012) | *MISQ*	R^2
Sykes (2015) | *MISQ*	R^2
Tan, Benbasat, & Cenfetelli (2013) | *MISQ*	f^2, R^2
Xue et al. (2013) | *JMIS*	Path coefficient significance
Wagner, Beimborn, & Weitzel (2014) | *JMIS*	f^2, R^2
Note: JAIS=Journal of the Association for Information Systems, JMIS=Journal of Management Information Systems, MISQ=MIS Quarterly	

Appendix B

The asymptotically efficient and consistent criteria described in this paper can be written as a function of the maximized value of the likelihood function ($\hat{L}$). For example,

$$AIC = -2ln\hat{L} + 2p_k$$

$$BIC = -2ln\hat{L} + p_k ln(n)$$

$$HQ = -2ln\hat{L} + 2p_k ln(ln(n))$$

Under a normal error distribution assumption, these likelihood-based formulas can be written in terms of SS_{error} as shown in Table B1 below (Burnham & Anderson, 2002; p.63; McQuarrie & Tsai, 1998).

Table B1. Information Theoretic Model Selection Criteria

Criterion	Formula	Description
Distance-based criteria		
Final prediction error (FPE)	$\left(\frac{SS_{error_k}}{n - p_k}\right) \times \left(1 + \frac{p_k}{n}\right)$	Selects the best model by minimizing the final prediction error.
Mallow's Cp	$\left(\frac{SS_{error_k}}{MS_{error}}\right) - (n - 2p_k)$	Based on mean square error (MSE); MS_{error} is MSE from the saturated (full) model.
Akaike information criterion (AIC)	$n\left[\log\left(\frac{SS_{error_k}}{n}\right) + \frac{2p_k}{n}\right]$	Estimates the relative expected KL distance to the unknown true model.
Unbiased AIC (AICu)	$n\left[\log\left(\frac{SS_{error_k}}{n - p_k}\right) + \frac{2p_k}{n}\right]$	Uses the unbiased estimate for population MSE, hence differs from AIC in small samples.
Corrected AIC (AICc)	$n\left[\log\left(\frac{SS_{error_k}}{n}\right) + \frac{n + p_k}{n - p_k - 2}\right]$	Corrects AIC's tendency to overfit (select a complicated model) under small samples.
Consistent criteria		
Bayesian information criterion (BIC)	$n\left[\log\left(\frac{SS_{error_k}}{n}\right) + \frac{p_k \log(n)}{n}\right]$	Derived using Bayesian argument; adjusts AIC for model complexity by using a stronger penalty for overfitting.
Geweke-Meese criterion (GM)	$\left(\frac{SS_{error_k}}{MS_{error}}\right) + p_k \log(n)$	Adjusts Mallow's Cp for model complexity by using a stronger penalty for overfitting.
Hannan-Quinn criterion (HQ)	$n\left[\log\left(\frac{SS_{error_k}}{n}\right) + \frac{2p_k \log(\log(n))}{n}\right]$	Corrects small sample performance of BIC by using a stronger penalty term.
Corrected HQ criterion (HQc)	$n\left[\log\left(\frac{SS_{error_k}}{n}\right) + \frac{2p_k \log(\log(n))}{n - p_k - 2}\right]$	Corrects small sample performance of HQ and adjusts for model complexity.

Note: $SSerror_{(k)}$ is the sum of squared errors for the k^{th} model in a set of models; MS_{error} is the mean squared error from the saturated model; p_k is the number of predictors in the k^{th} model plus 1.

To compute the criteria in Table B1, researchers can calculate the SS_{error} for the partial regression of a certain target construct as follows:

$$SS_{error} = (1 - R^2)SS_{total},$$

where R^2 is the coefficient of determination of the target construct for the k^{th} model in the competing set. As the PLS algorithm uses standardized latent variable scores as input for the partial regressions in the structural model, SS_{total} is equal to N-1. Hence,

$$SS_{error} = (1 - R^2)(N - 1)$$

To calculate MS_{error}, researchers need to calculate the R^2 of the saturated model in which all constructs are linked to the target construct:

$$MS_{error} = (1 - R^2_{saturated})$$

Finally, p_k is the number of immediate antecedents predicting the target construct plus 1.

We illustrate the computation of all criteria in an Excel sheet, which can be obtained from the Downloads section at https://www.pls-sem.net/.

About the Authors

Pratyush Nidhi Sharma is an assistant professor in the Alfred Lerner College of Business & Economics, University of Delaware. He received his PhD from the University of Pittsburgh in 2013. His research interests include online collaboration communities and networks, open source software development, and research methods used in information systems, particularly partial least squares path modeling. In addition, he is also interested in issues surrounding technology use and adoption and human computer interaction. His research has been published in distinguished journals such as the *Journal of Retailing*, *Decision Sciences*, *Government Information Quarterly*, *Journal of Information Systems*, *Journal of Business Research*, *Journal of International Marketing*, and *International Journal of Accounting Information Systems*. In addition, he has published several book chapters and presented his research at premier conferences such as the International Conference on Information Systems, Americas Conference on Information Systems, INFORMS, and PLS.

Marko Sarstedt is a chaired professor of marketing at the Otto-von-Guericke-University Magdeburg (Germany) and adjunct professor at Monash University Malaysia. His main research interest is the advancement of research methods to further the understanding of consumer behavior. His research has been published in *Journal of Marketing Research, Journal of the Academy of Marketing Science, Multivariate Behavioral Research, Organizational Research Methods, MIS Quarterly, Decision Sciences, Journal of World Business,* and *Long Range Planning*, among others. Marko has coedited several special issues of leading journals and co-authored four widely adopted textbooks, including *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (together with Joe F. Hair, G. Tomas M. Hult, and Christian M. Ringle). Marko has been named a member of Clarivate Analytic's 2018 Highly Cited Researcher list.

Galit Shmueli is the Tsing Hua Distinguished Professor at the Institute of Service Science, National Tsing Hua University, Taiwan. She is also the director of the Center for Service Innovation & Analytics at NTHU's College of Technology Management. She earned her Master's and PhD degrees in statistics from the Technion-Israel Institute of Technology and her BA in psychology and statistics from Haifa University. Her research focuses on statistical and data mining methodology with applications in information systems and healthcare and a focus on prediction vs. explanation. She has authored and co-authored multiple books and publications in peer-reviewed journals, including *Management Science, Journal of the American Statistical Association, Journal of the Royal Statistical Society, Information Systems Research, MIS Quarterly, Journal of the Association of Information Systems*, *Journal of Business and Economic Statistics, Marketing Science, Statistical Science, Technometrics,* and *Proceedings of the National Academies of Science.*

Kevin H. Kim (†) was an associate professor of education in the Department of Psychology at the University of Pittsburgh. He passed away in July 2014. His scholarship was in statistical methods as well as in the behavioral and social sciences. His work focused on structural equation modeling, multivariate statistics, multilevel modeling, and Asian American mental health. He published extensively in methodology and in several other academic disciplines and worked as a co-investigator on numerous federally funded grants with faculty throughout Pitt and other universities. In 2007, he cowrote the book titled *Univariate and Multivariate General Linear Models: Theory and Applications.* He was also a highly regarded teacher of statistical methodology at Pitt, teaching numerous graduate-level courses and supervising many PhD students.

Kai Oliver Thiele received his PhD from the Hamburg University of Technology in Germany. He holds an MBA degree from the University of Kansas and a Master's degree in business mathematics from the University of Hamburg. He received a DAAD scholarship and a scholarship from the German Institute for Japanese Studies to conduct his research in Japan.