

Rigdon, Edward; Sarstedt, Marko; Moisescu, Ovidiu-loan

Article — Published Version

Quantifying model selection uncertainty via bootstrapping and Akaike weights

International Journal of Consumer Studies

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Rigdon, Edward; Sarstedt, Marko; Moisescu, Ovidiu-loan (2023) : Quantifying model selection uncertainty via bootstrapping and Akaike weights, International Journal of Consumer Studies, ISSN 1470-6431, Wiley, Hoboken, NJ, Vol. 47, Iss. 4, pp. 1596-1608,
<https://doi.org/10.1111/ijcs.12906>

This Version is available at:

<https://hdl.handle.net/10419/287960>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Quantifying model selection uncertainty via bootstrapping and Akaike weights

Edward Rigdon¹  | Marko Sarstedt^{2,3}  | Ovidiu-Ioan Moisescu³ 

¹Marketing Department, Georgia State University, Atlanta, Georgia, USA

²LMU Munich School of Management, Ludwig-Maximilians-University Munich, Munich, Germany

³Faculty of Economics and Business Administration, Marketing Department, Babeş-Bolyai University, Cluj-Napoca, Romania

Correspondence

Marko Sarstedt, Ludwig-Maximilians-University Munich, Munich School of Management, Ludwigstr. 28 RG/III, 80539 Munich, Germany.
Email: sarstedt@lmu.de

Funding information

Ludwig-Maximilians-Universitaet Munchen

Abstract

Picking one ‘winner’ model for researching a certain phenomenon while discarding the rest implies a confidence that may misrepresent the evidence. Multimodel inference allows researchers to more accurately represent their uncertainty about which model is ‘best’. But multimodel inference, with Akaike weights—weights reflecting the relative probability of each candidate model—and bootstrapping, can also be used to quantify model selection uncertainty, in the form of empirical variation in parameter estimates across models, while minimizing bias from dubious assumptions. This paper describes this approach. Results from a simulation example and an empirical study on the impact of perceived brand environmental responsibility on customer loyalty illustrate and provide support for our proposed approach.

KEYWORDS

Akaike weights, bootstrapping, information criteria, model selection, uncertainty

1 | INTRODUCTION

The mechanisms underlying consumer behaviour are inherently complex due to the multitude of causal interactions and processes. Consumer researchers draw on theory and logic to derive models that offer partial explanations of these mechanisms. However, such models are always derived from specific theoretical lenses and reflect different assumptions, interpretations, and levels of development (Sharma et al., 2019). As a result, there is not a single best model that explains a certain mechanism well, but rather a set of models which all account for the observed phenomena in a reasonable way. Researchers therefore routinely estimate different models and pick the ‘best’ model based on model fit statistics and conceptual considerations (Preacher & Merkle, 2012)—albeit this process is rarely made transparent (e.g., John et al., 2012). Such multimodel inference recognizes the practical reality that (1) researchers often face difficult choices among multiple competing hypotheses rather than just two mutually exclusive possibilities, and (2) the choice among the competitors is often

not clear-cut, so that researchers have an evidence-based reluctance to discard all but one.

However, choosing a single ‘best’ model means ignoring all of the evidence that favoured other models, not to mention any evidence that was inconclusive (Danks et al., 2020). In an extreme case with only two models in the set, 51% of the evidence could be translated into 100% confidence in one model versus the other. This practice ignores the uncertainty inherent in the model selection process (Preacher & Merkle, 2012).

The problem of model selection uncertainty has been acknowledged in research in several fields of scientific inquiry. For example, Burnham and Anderson (2002) present a variety of methods that can be used to measure the uncertainty associated with model selection. Preacher and Merkle (2012) show that model selection uncertainty can be considerable under common circumstances, leading to overconfidence in the selected model. Finally, Lubke and Campbell (2016) use bootstrapping to compute a selection rate that guides model selection and safeguards the model's replicability. While these studies

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *International Journal of Consumer Studies* published by John Wiley & Sons Ltd.

have advanced our understanding of model selection uncertainty, they rely on an error-centric framework in that the authors define uncertainty as a function of sampling variance only. However, as we will conceptually argue and empirically show, sampling variance is only one element of the uncertainty that comes with any model selection.

Capturing the other elements of model selection uncertainty requires reaching beyond concepts discussed in the social sciences and consider ‘uncertainty’ as done in metrology—measurement science in the physical sciences (Bell, 1999). In metrology, uncertainty is defined as quantified doubt about the value of the measurand, which is the quantity whose value is sought (JCGM, 2012). High uncertainty implies that a measurement is consistent with a wide range of plausible values for the measurand, which may lead to a wide range of obtained values across different measurements and studies (Rigdon et al., 2020). Researchers in the physical sciences not only quantify overall uncertainty but also catalogue the material components or sources of uncertainty, with the aim of perhaps improving design, instrumentation, and execution in order to further reduce uncertainty in future research (Bell, 1999). Those material components include potential doubt about which statistical model is ‘best’.

While standard in the physical sciences, the concept of uncertainty is largely unknown to consumer researchers. This is surprising considering that consumer research focuses on perceptions, intentions, and behaviours that are inherently difficult to measure. For example, ambiguous concept definitions, item selection, and measurement validation contribute to uncertainty in consumer research studies (Rigdon & Sarstedt, 2022). Researchers have therefore started calling for bringing an uncertainty-centric research framework to consumer research (Bach et al., 2020; Fisher & Stenner, 2017; Rigdon et al., 2020), emphasizing that such an undertaking requires considerable effort (Rigdon & Sarstedt, 2022).¹

The aim of this paper is to take a step in that direction by describing an approach for quantifying model selection uncertainty that captures variation in parameter estimates across models. Our approach extends beyond prior research in that it captures all material components that emerge from the model selection process—rather than focusing on sampling variance only. To provide support for and illustrate our approach, we present the results of a simulation study and an empirical study on the impact of perceived brand environmental responsibility on customer loyalty. By quantifying model selection uncertainty in full, our approach offers the means to contextualize research findings and improve model selection practice over time.

2 | QUANTIFYING MODEL SELECTION UNCERTAINTY

Model selection uncertainty exists whenever researchers could plausibly consider alternative statistical models that imply different values for parameter estimates, thereby increasing the range of possible estimates which might plausibly be observed and, thus, might apply in the real world. Assume a fixed effects model, where parameters have fixed but unknown values. The total variance of a parameter estimate

can be represented as the sum of bias squared plus random sampling variance (Fox, 1997; Kennedy, 2008). In a conventional analysis, if an estimator is unbiased given certain supporting assumptions, and if all of those supporting assumptions held exactly, then the variance in a parameter estimate would by definition be purely a matter of random sampling variance (e.g., Cohen et al., 2003).

Those supporting assumptions would typically include the assumption that the model containing the parameter is correct in the population. This assumption, in particular, is untenable in situations with multiple candidate models, since the models cannot all be correct at the same time. More likely, none of the models is anything more than a reasonable approximation (Burnham & Anderson, 2002; Sweeten, 2020). Because of this, metrics that depend on the model being correct are potentially misleading. This could include model test statistics, such as the chi-square statistic in factor-based structural equation modelling and related metrics that quantify the divergence between the covariance matrix as implied by the model specified by the researcher—which is assumed to be correct in the population—and the observed or empirical covariance matrix. Similarly, estimated standard errors provided by statistical software such as AMOS, LISREL, and Mplus are unlikely to reflect the uncertainty that comes with the model selection task. Hence, any approach to quantify model selection uncertainty needs to avoid relying upon those standard statistics. However, model selection uncertainty can be quantified, without relying on statistics that fail to capture model selection uncertainty's full scope. By relying on a combination of information-theoretic model selection criteria and nonparametric bootstrapping, the approach relaxes the assumption of one model being correct in the population, while quantifying the variability of parameter estimates that goes beyond mere sampling variance.

Akaike (1973) is credited with introducing and popularizing the idea of approaching the relative probability of candidate models through the likelihood of each model given the data, using what is called the Akaike information criterion (AIC). The AIC is the most prominent among an assortment of information-theoretic criteria that seek to select the model being closest to the unknown true model which generates the data.² Being a relative fit measure, the AIC expresses the relative distances of competing models from the unknown true model, with the ‘best’ model having the smallest AIC value. In more formal terms, given a data set containing n ($n = 1, \dots, N$) observations on p variables, a model m can be defined by a vector of k parameter estimates $\hat{\theta}$. If the data are normally distributed, then $\log \mathcal{L}(\hat{\theta} | \text{data})$, the log likelihood of the data given the estimated parameters, is a biased estimate of the Kullback–Leibler information of the model, with the bias approximated by k (Burnham & Anderson, 2002). Akaike (1973) derived the formula

$$AIC = -2 \log \mathcal{L}(\hat{\theta} | \text{data}) + 2k. \quad (1)$$

Smaller values in AIC indicate a superior model in information-theoretic terms.³ The derivation behind AIC requires only that the candidate models be reasonably close to the true model in terms of

their Kullback–Leibler distance (Claeskens & Hjort, 2008, Chap. 2). None of the candidate models need be strictly ‘correct’ (Preacher & Merkle, 2012).

While the AIC proves useful for ranking and selecting models, the models’ actual criterion values do not offer any insights regarding the relative weights of evidence in favour of models under consideration (Danks et al., 2020). This particularly holds since, in empirical applications, the AIC differences are typically small (Preacher & Merkle, 2012). To address this concern, Akaike (1983) proposed computing Akaike weights that draw on the criterion’s values to express each model’s relative strength of evidence *vis-à-vis* the alternative models (Wagenmakers & Farrell, 2004). With values AIC_i for all candidate models m_i , $i = 1$ to l , the process of deriving weights reflecting relative probabilities begins by finding $\Delta_i = AIC_i - \min(AIC)$, the minimum AIC value across the set of candidates. The relative likelihood of each candidate, relative to the set, is then (Danks et al., 2020, p. 15, Equation 2)

$$L(m_i) \propto \exp\left(-\frac{1}{2}\Delta_i\right), \quad (2)$$

where the symbol \propto means ‘proportional to’. These relative likelihoods can be transformed into probabilities or weights w_i (Danks et al., 2020, p. 15, Equation 3)

$$w_i = \frac{L(m_i)}{\sum_i L(m_i)}, \quad (3)$$

with $\sum_{i=1}^l w_i = 1$.

To estimate the actual uncertainty associated with a parameter estimate, given multiple plausible models, a researcher might consider taking the sum of the parameter estimates’ variances (squares of the estimated standard errors), weighted by the Akaike weights. But because the estimated standard errors carry the assumption that the estimated model is correct, the resulting weighted average would be not adequately reflect the model selection-induced uncertainty.

To estimate model selection uncertainty without relying on this ‘model is correct’ assumption, researchers can draw on a (nonparametric) bootstrapping procedure. In bootstrapping, a great number of samples is drawn from the original data set with replacement to generate an empirical distribution for the model estimates (Efron & Tibshirani, 1986). The empirical distribution facilitates computing the variance and standard error (i.e., the standard deviation of the estimates) for inference testing. This variance estimate incorporates both random sampling variance *and* model selection uncertainty. To capture the variability of a parameter estimate across the different candidate models, researchers need to apply the following steps:

1. Calculate Akaike weight w_i for each candidate model m_i .
2. Choose a total number of bootstrap samples R . A large number of samples such as 10,000 minimizes rounding when translating model probabilities into numbers of bootstrap resamples, but requires more computational time.

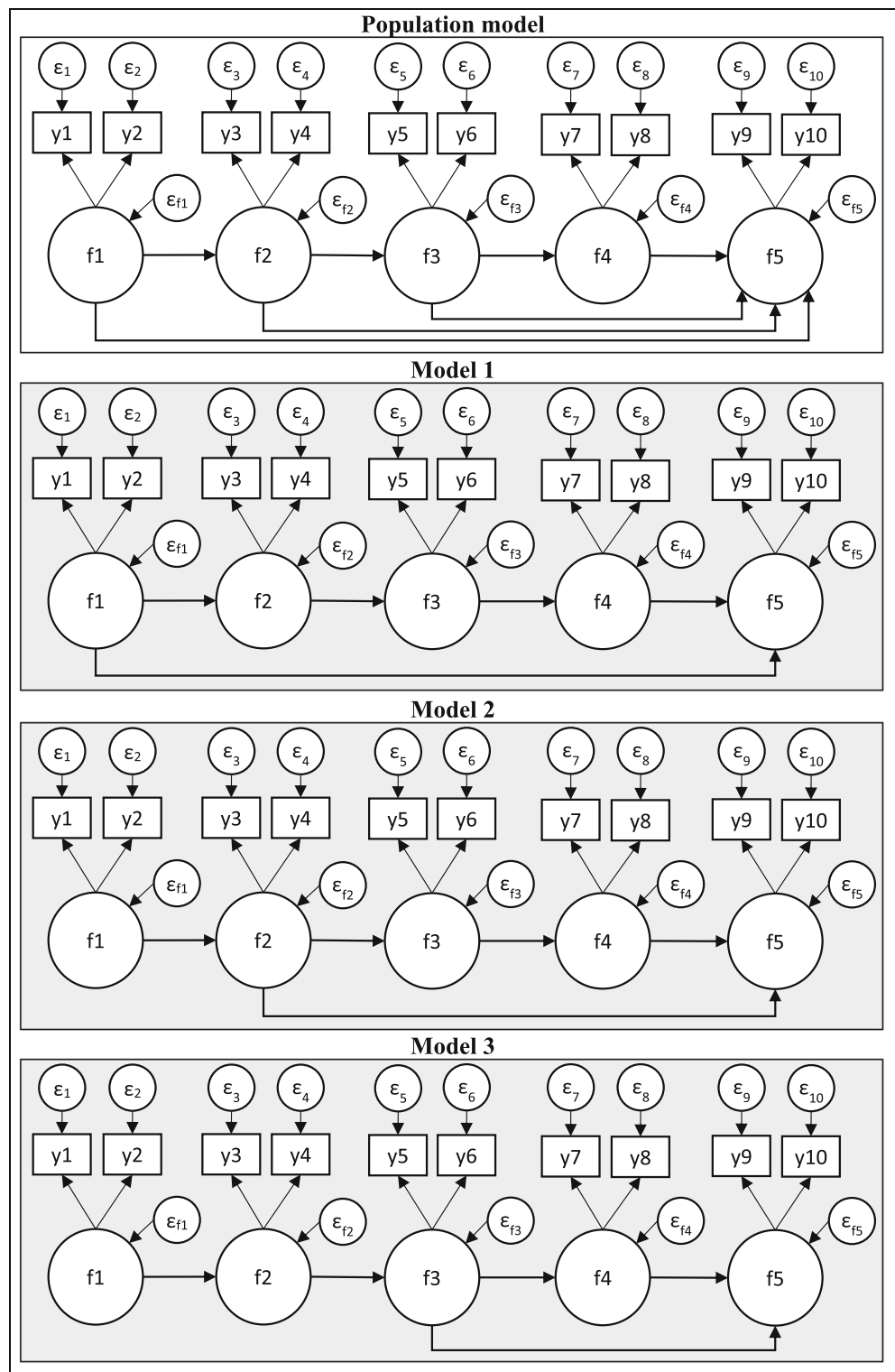
3. For each model m_i , draw $R_i = R \cdot w_i$ bootstrap samples and estimate the parameter within that model. For example, if a researcher considers two models with equal Akaike weights and 10,000 total bootstrap samples, then the study will draw $10,000 \times 0.5 = 5000$ bootstrap samples for each model and estimate each model’s parameters $R_i = 5000$ times. The number of bootstrap samples must be large enough to meet each model’s minimum sample size requirement.
4. Combine the R_i estimates of the parameter from each model and compute the standard deviation—a standard error that reflects both uncertainty due to random sampling variance and model selection uncertainty, as it captures the variability of the parameter estimate across the different candidate models.
5. Calculate an overall uncertainty interval for the parameter estimate across all candidate models and compare it with the intervals from the individual models.

3 | SIMULATION EXAMPLE

In the following, we offer a simulation study to quantify the uncertainty that comes with a model selection task. To do so, we compute an overall uncertainty interval for a parameter estimate across several models, and contrast this interval with those derived from the single models. A wider combined interval suggests additional variance introduced through the model selection.

We created a data set of 500 observations for the purposes of estimating the parameters of three common factor models shown in Figure 1. We chose the simplex model (a linear sequence of effects) as a simple baseline (Jöreskog, 1970). The simplex model features prominently in psychological research, where it is commonly used to measure constructs repeatedly on the same consumers over several occasion (Hayduk, 1994). The three simplex models include five common factors and ten observed variables, but differ in the structural paths included. All three models include paths from factor 1 to factor 2, from factor 2 to factor 3, and so forth. We chose this model type as its simple structures offers a suitable basis for illustrating our approach’s basic principle. Candidate model 1 includes an additional direct path from factor 1 to factor 5. Candidate model 2 includes an additional direct path from factor 2 to factor 5, and candidate model 3 includes an additional direct path from factor 3 to factor 5. The population, from which the data were sampled, included all three additional paths, so that none of the estimated models is entirely correct in the population. In line with simulation studies in structural equation modelling (Cho et al., 2022; Hair et al., 2017; Sarstedt et al., 2016), we selected pre-specified path coefficients of 0.5 for the first three relationships of the simplex model (i.e., $\beta_{12} = \beta_{23} = \beta_{34} = 0.5$) and 0.4 for the final relationship (i.e., $\beta_{45} = 0.4$)—the last path uses 0.4 so that the additional direct paths can potentially explain a significant share of the terminal factor’s variance. The additional paths of models 1–3 were set to $\beta_{15} = 0.2$, $\beta_{25} = 0.2$, and $\beta_{35} = -0.3$. The indicator loadings were set to 0.8 for the first and 0.7 for the second indicator. These values were chosen to safeguard the measures convergent validity

FIGURE 1 Simulation models.



and internal consistency reliability, which are functions of the indicator loadings. The parameter of interest here is the path coefficient β_{45} . Given the direct paths to the terminal factor which are excluded from various candidate models, this parameter should demonstrate a substantial degree of variability across these models.

Multivariate normal data were randomly generated and factor models estimated using the R environment (R Core Team, 2022)

and version 0.6.8 of the lavaan package (Rosseel, 2012). Appendix A includes the covariance matrix of the observed variables for the data set used. Appendix B includes a complete R script.

The results in Table 1 show that the fit statistics for the three candidate models are very similar. All three models have χ^2 values which, given 30 degrees of freedom, imply p values less than .001 and

TABLE 1 Results for the three models.

	Model 1	Model 2	Model 3
Parameter estimate	0.289	0.262	0.455
Estimated standard error	0.061	0.062	0.079
χ^2 (df = 30, p value < .001)	68.625	71.281	72.033
CFI	0.974	0.972	0.972
RMSEA	0.051	0.052	0.053
AIC	13,340.479	13,343.135	13,343.887
Δ	0	2.656	3.408
Exp ($-1/2 \times \Delta$)	1	0.265	0.182
Akaike weight	0.691	0.183	0.126

thus lead to rejection of the hypothesis of exact fit. All three models have CFI and RMSEA values that can be taken to indicate good approximate fit (Hu & Bentler, 1999). Hence, each of these models can be judged to have good approximate fit if considered by itself. Recall that in this simulation context, none of the three estimated models is equivalent to the model that generated the data; in practice, of course, researchers do not have this information.

Estimates of the focal path coefficient β_{45} vary somewhat across the three candidate models. The estimates range from 0.262 (Model 2) to 0.455 (Model 3). Estimated standard errors range from 0.061 (Model 1) to 0.079 (Model 3).

AIC values for the three candidate models were calculated using Equation (1). Equations (2) and (3) were used to calculate Akaike weights (Table 1). A researcher wishing to select the 'best' model based on the Akaike weights would therefore clearly favour Model 1 (0.691) over Model 2 and Model 3, which yield weights of 0.183 and 0.126, respectively, but choosing Model 1 as the exclusive 'winner' would mean ignoring the 31% of the evidence favouring the other models. Multiplying the Akaike weights times 10,000 bootstrap samples provides the number of samples to be bootstrapped for each candidate model (e.g., for Model 1: $R_i = 0.691 \times 10,000 = 6910$ bootstrap samples). Next, we ran bootstrapping with the indicated number of resamples assigned to each candidate model, using the original data set, and computed the parameter estimates for each bootstrap run.⁴ This provided the data to calculate the empirical standard deviation of the parameter estimates—that is, the standard error—weighted by the Akaike weights.

Across 10,000 total bootstrap samples, with each candidate model weighted by the Akaike weights, the estimated mean path coefficient β_{45} is 0.310 with a standard error of 0.098. At the same time, however, the estimated standard error for this parameter, associated with each separate model mirrored the original results, ranging from 0.061 (Model 1) to 0.079 (Model 3). That is, the bootstrap value, which takes into account the *multiple* candidate models, demonstrates more variability, hence more uncertainty, in the estimate of the focal path coefficient.

These results provide the basis for calculating an overall uncertainty interval for the parameter estimate across the three models together, which can be compared to confidence intervals derived from the original single-model parameter estimates and standard errors.

Sample size is large, so in all cases an approximately 95% interval can be constructed as estimated mean $\pm 1.96 \times$ standard error, which yields the following intervals (Figure 2):

Model 1: (0.169, 0.409)
 Model 2: (0.140, 0.384)
 Model 3: (0.300, 0.610)
 Combined: (0.118, 0.502)

The 95% uncertainty interval, accounting for model selection uncertainty and random sampling variance instead of only random sampling variance, is wider and includes values closer to 0. The combined result alerts the researcher that outcomes from a policy implementation may be more varied than they would otherwise expect. Researchers designing a future study might be surprised to find that a planned sample size yields lower statistical power than they had been counting on—or higher power, in which case money spent on sample size may have been wasted. Competing conclusions across different studies may actually all be consistent with this study's results, granting the wider range of uncertainty implied by this approach. In this study, even the expanded uncertainty range excluded 0. In other situations, the impact of accounting for model selection uncertainty could be to keep alive the possibility that there is no non-zero effect at all.

In this example, recall that the Akaike weight for Model 1 was $w_1 = 0.691$; that is, almost 70%, dominating the 10,000 bootstrap samples. In a situation with several very competitive candidate models, the impact of model selection uncertainty could be substantially larger. And, again, this approach only accounts for random sampling variance and model selection uncertainty. In this simulation-based scenario, other potential sources of uncertainty can be discounted. In research with real data, researchers would likely need to account for several other material sources of uncertainty.

Note that attempting to create a weighted standard error—by squaring the reported standard errors for the three models, weighting them using the Akaike weights, then taking the square root to create a weighted standard error—necessarily fails. Even in this situation, where distributional assumptions underlying maximum likelihood estimation hold, the standard errors for each candidate model are ignorant of the existence of other candidate models, and so can hardly account for the impact of model selection uncertainty. Moreover, as noted previously, the existence of multiple plausible candidate models undermines the assumption that any one candidate model is correct, even though all three candidate models here show what could be called good approximate fit. Of course, bootstrapped standard errors and standard errors derived from distributional assumptions need not be equal, certainly not in any limited sample (Efron & Tibshirani, 1986). In circumstances where distributional assumptions are avoided, as is typical with composite-based methods like partial least squares path modelling (Lohmöller, 1989) or generalized structured component analysis (Hwang & Takane, 2004), this procedure illustrates a very plausible approach.

FIGURE 2 Confidence/uncertainty intervals for candidate models and combined model.

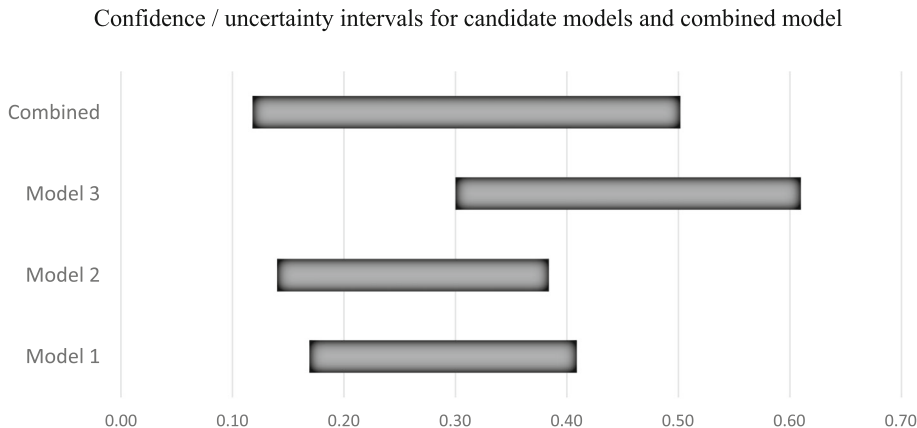
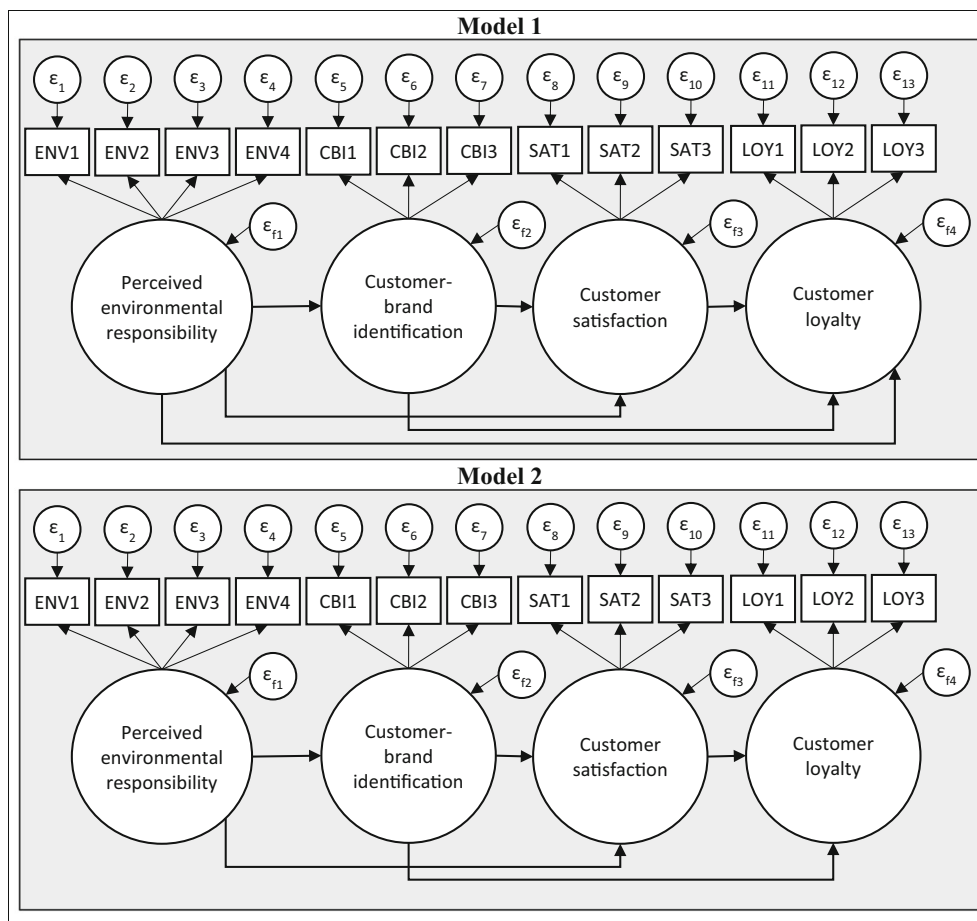


FIGURE 3 Empirical models.



4 | EMPIRICAL EXAMPLE

To illustrate our approach for addressing model selection uncertainty, we used a model depicting the impact of perceived brand environmental responsibility—the extent to which stakeholders view a brand as accounting for environmental interests (Öberseder et al., 2014)—on customer loyalty, via customer-brand identification (Bhattacharya & Sen, 2003) and customer satisfaction. While most of the model relationships are well established, prior research is inconclusive regarding the effect of perceived brand environmental responsibility on customer loyalty. Specifically, some studies suggest that this impact is fully mediated by customer satisfaction (He & Li, 2011; Pérez & Del

Bosque, 2015), whereas others indicate a direct influence on customer loyalty (Du et al., 2007; Glaveli, 2021). Hence, we proposed two competing models, as depicted in Figure 3.

To measure the four constructs, we adapted scales from Öberse-der et al. (2014) and Turker (2009) for perceived brand environmental responsibility, from Mael and Ashforth (1992) for customer-brand identification, from Cronin et al. (2000) for customer satisfaction, and from Zeithalm et al. (1996) for customer loyalty. Table C1 in Appendix C reports the item wordings along with validity and reliability estimates.

The analysis draws on data collected by means of a survey conducted among customers of personal care brands from Romania using

quota sampling. The final data set comprises answers from 1167 customers (50.6% men, and 49.4% women; 37.7% under 30 years, 35.3% between 30 and 44 years, and 27.0% older).⁵ We use the SEMinR package (Ray et al., 2022) for the R statistical environment (R Core Team, 2022) for model estimation.

Model-fit measures for the proposed four-factor model meet the generally suggested thresholds (Hu & Bentler, 1999), indicating good approximate but not exact fit ($\chi^2 = 278.991$, $df = 59$, $p = .000$; CFI = 0.974 > 0.95, SRMR = 0.035 < 0.08, and RMSEA = 0.057 < 0.06). Furthermore, the analysis offers support for measures' internal consistency reliability, convergent validity, and discriminant validity (Tables C1 and C2 in Appendix C).

The results in Table 2 show that the two competing models have very similar fit statistics. Neither of them achieves exact fit (χ^2 with p values less than .001), but both have CFI, SRMR and RMSEA values that indicate good approximate fit (Hu & Bentler, 1999). The results

of the AIC computation clearly speak in favour of Model 1, as indicated by its considerably larger Akaike weight (0.837 vs. 0.163).

Following the procedure outlined above, we computed 10,000 bootstrap samples and, using the Akaike weights (Table 2), determined the number of samples to be bootstrapped for each candidate model (i.e., 8370 for Model 1, and 1630 for Model 2). Next, we ran bootstrapping with the indicated number of samples for each model, using the original data set, and computed the parameter estimates for each bootstrap run. Eventually, we combined the parameter estimates of all 10,000 bootstrap runs, and computed bootstrap estimates, standard errors, and 95% intervals for all model relationships (Table 3).

According to the proposed procedure, model selection uncertainty is reflected by the discrepancies between (1) the standard errors and confidence intervals of the candidate models' parameter estimates, computed based on 8370 bootstrap samples for Model 1, and 1630 for Model 2, and (2) the standard errors and uncertainty intervals computed by combining parameter estimates of all 10,000 bootstrap samples. Additionally, the discrepancy between standard errors derived from bootstrap samples and the original standard errors also reflect uncertainty related to the distributional assumptions supporting the original parametric estimation method. Comparing the results for the various models, we find that the estimates are very stable. For example, standard errors and confidence intervals for the ENV → CBI relationship are very similar across Model 1, Model 2, and the combined sample. The combined bootstrap standard error for the CBI → LOY relationship (0.033) demonstrates slightly more variability than the bootstrap standard error in Model 1 (0.032). Comparing the combined confidence interval (0.070, 0.200) to that of Model 1 (0.068, 0.195) suggests that selecting Model 1 may lead to a small degree of overconfidence regarding the direct effect of customer-brand identification on customer loyalty. Overall, however, these results indicate that the model selection task does not induce a substantial degree of additional uncertainty.

TABLE 2 Structural models assessment.

	Model 1	Model 2
Parameter estimates		
ENV → CBI	0.370***	0.371***
CBI → SAT	0.211***	0.207***
ENV → SAT	0.233***	0.243***
CBI → LOY	0.131***	0.152***
SAT → LOY	0.720***	0.735***
ENV → LOY	0.063*	N.A.
Model fit		
χ^2	278.991*** (df = 59)	284.258*** (df = 60)
CFI	0.974	0.974
SRMR	0.035	0.038
RMSEA	0.057	0.057
Model selection		
AIC	342.991	346.258
Δ	0	3.267
Exp $(-1/2 \times \Delta)$	1	0.195
Akaike weights	0.837	0.163

*** $p < .001$. * $p < .05$.

5 | CONCLUSION

5.1 | Discussion and implications

This paper extends the pioneering work of Burnham and Anderson (2002, 2004) by proposing a procedure for quantifying uncertainty elements in

TABLE 3 Standard errors and confidence/uncertainty intervals.

	Model 1 (8370 samples)				Model 2 (1630 samples)				Combined (10,000 samples)			
	B	SE	LB	UB	B	SE	LB	UB	B	SE	LB	UB
ENV → CBI	0.369	0.032	0.307	0.431	0.372	0.031	0.311	0.434	0.369	0.032	0.307	0.431
CBI → SAT	0.209	0.039	0.134	0.285	0.204	0.038	0.129	0.280	0.209	0.039	0.133	0.285
ENV → SAT	0.234	0.036	0.162	0.305	0.243	0.036	0.172	0.315	0.235	0.037	0.164	0.307
CBI → LOY	0.131	0.032	0.068	0.195	0.153	0.030	0.093	0.212	0.135	0.033	0.070	0.200
SAT → LOY	0.720	0.027	0.668	0.772	0.734	0.025	0.685	0.783	0.722	0.027	0.670	0.775
ENV → LOY	0.063	0.034	-0.004	0.129	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Note: B, bootstrap parameter estimate; SE, bootstrap estimated standard error; LB/UB, 95% interval lower/upper bound (computed as $B \pm 1.96 \times SE$).

multimodel inference beyond mere sampling variance. There are several implications that follow not just from this specific procedure but also from what it suggests about the impact of researchers' model comparison practice on the uncertainty of the results obtained. Uncertainty exists whether it is quantified and documented or not. Standard errors for a single candidate model implicitly rest on the assumption that a single candidate model is being tested against a null hypothesis—regardless whether the models are theoretically motivated or whether they have been derived from an exploratory perspective (Liengard et al., 2021). If researchers consider multiple models, then it is quite likely that parameter estimates vary across models. In terms of researchers' uncertainty about the actual value of the parameter in the population, the existence of multiple candidate models cannot be ignored.

The ASA has taken great pains to overcome misunderstandings about p values (Wasserstein et al., 2019), and researchers in multiple fields are increasingly concerned about trying to root out statistical procedures being abused to produce the p value that makes research publishable (Hussey, 2021). These abuses include researchers estimating many models but then reporting only the one that supports a maximum share of hypothesized relationships. The problem is compounded if the many candidate models arise from opportunistic exploration or 'fishing' rather than from divergent a priori theoretical perspectives. Estimating many models carries costs, including in terms of increased uncertainty. If a cost is never quantified, then it will never appropriately influence decision-making. This makes it very important that researchers begin quantifying all aspects of uncertainty related to data analysis procedure (Rigdon et al., 2020).

The most ominous implication of this research is that it accounts for only one small component of the overall uncertainty that consumer researchers generally ignore. The effect of accounting for model selection uncertainty, in this hypothetical example, is to expand the *ex post* range of uncertainty regarding a given measurand. Incorporating additional material components of uncertainty will only further widen the resulting window of doubt, possibly leading to the conclusion that many or most studies are producing no knowledge at all. But part of the point of building uncertainty budgets is that researchers become aware of sources of uncertainty that need to be controlled. Including such controls in future research designs will lead to studies that more surely produce actual contributions to the body of consumer research knowledge.

5.2 | Limitations and future research

A primary limitation of this approach to quantifying model selection uncertainty is that it assumes that the models involved were all chosen a priori, independent of the data. An alternative approach, probably quite common, is that a researcher develops candidate models in response to the data (Danilov & Magnus, 2004; Simmons et al., 2011), favouring those that appear to be consistent with the data at hand. In this alternative situation, an assessment of uncertainty must encompass not only the parameter estimates arising from the set of models that the researchers actually considered but also the parameter

estimates that *would have arisen* from the other models that the researchers might have considered if the researchers had observed different data (Gelman & Loken, 2014).

One caution worth noting is that the behaviour of AIC and related information-theoretic model selection criteria can be sensitive to sample size. In cases where the number of parameters being estimated is large relative to sample size, an adjusted AIC with a different penalty function may be appropriate (Burnham & Anderson, 2004) such as Bedrick and Tsai's (1994) AIC_c. Similarly, future research should consider asymptotically consistent information theoretic criteria such as the BIC and Geweke and Meese's (1981) criterion, which have been shown to perform well in structural equation modelling-based model selection tasks (Sharma et al., 2019, 2021). Such an assessment requires comparing models with different degrees of freedom in order to impose different penalty functions on the candidate models.

Finally, we have demonstrated the uncertainty that comes with a model selection task in the context of common factor-based methods like covariance structure analysis where the Akaike weights express a model's overall fit, relative to the other candidate models in the set. However, Akaike weights can also be computed in the context of composite-based methods like partial least squares path modelling or generalized structured component analysis (Sharma et al., 2019) where they express a model's relative fit with regard to a specific endogenous construct. Such an assessment leaves the model selection uncertainty triggered by differences in relationships in other parts of the model largely untapped. Nevertheless, future research should consider composite-based SEM method's performance in that regard, focusing on smaller models and acknowledging the limitations of focusing on a specific endogenous construct.

5.3 | Concluding remarks

The approach described here effectively inflates the assessment of uncertainty due strictly to random sampling variance to also include model selection uncertainty. Unfortunately, that leaves additional, potentially material sources of uncertainty that impact other aspects of the research practice unaccounted for. It will probably be a while before quantitative social science has anything close to a full set of tools for quantifying uncertainty. Lacking a full assessment of uncertainty, researchers need to acknowledge that any evaluation of uncertainty strictly on the basis of statistical p values—as done in common accounts of uncertainty (Preacher & Merkle, 2012) is not enough. This call goes beyond the ASA's concerns regarding 'statistical significance' as a criterion for judging the meaning and worth of research findings. To the extent that research is evaluated in terms of relevance—the relationship of research fundings to real-world problems—consumer researchers need to acknowledge the full extent of uncertainty associated with research fundings. Doing so will increase confidence in and practical relevance of consumer research findings.

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

WOA Institution: Ludwig-Maximilians-Universitaet Munchen.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at: https://osf.io/usymf/?view_only=11dd4d4975ba4705bfb152e68ca25ff8. The simulation code is included in the Appendix.

ORCID

Edward Rigdon  <https://orcid.org/0000-0003-4904-3952>

Marko Sarstedt  <https://orcid.org/0000-0002-5424-4268>

Ovidiu-Ioan Moisescu  <https://orcid.org/0000-0003-0097-5660>

ENDNOTES

¹ Note, however, that some fields of consumer research also rely on laboratory instruments that allow for a more straightforward quantification of uncertainty (e.g., functional magnetic resonance imaging, electroencephalography, and eye tracking; Camerer & Yoon, 2015; de Jong & Pieters, 2019; Oliveira et al., 2023).

² Criteria in another stream such as Schwarz (1978) Bayesian information criterion, provide estimates of the posterior probability of a model being true with researchers choosing the model that maximizes this probability on a given data set. However, as these criteria are based on the same principle—an empirical likelihood with a penalty that favours sparse models over more saturated models—we focus on the AIC as the primary criterion.

³ The AIC is typically written as a function of the maximum value of the likelihood function, but the metric can be derived from the average squared residual across observations (Burnham & Anderson, 2002, p. 62), provided that the error distribution is normal with a constant variance. This characteristic makes the AIC amenable to regression-based techniques such as partial least squares path modelling (Sharma et al., 2019) and generalized structured component analysis (Hwang & Takane, 2004), which do not rely on maximum likelihood estimation (Hair et al., 2022).

⁴ Convergence failed on 16 of the original bootstrap samples—11 for Model 1, 4 for Model 2 and 1 for Model 3. We reran the bootstrapping procedure to generate additional samples for each model.

⁵ The dataset can be downloaded from the OSF at https://osf.io/usymf/?view_only=None.

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). Springer.
- Akaike, H. (1983). Information measures and model selection. *International Statistical Institute*, 44, 277–291.
- Bach, D. R., Melinscak, F., Fleming, S. M., & Voelke, M. C. (2020). Calibrating the experimental measurement of psychological attributes. *Nature Human Behaviour*, 4(12), 1229–1235. <https://doi.org/10.1038/s41562-020-00976-8>
- Bedrick, E. J., & Tsai, C. L. (1994). Model selection for multivariate regression in small samples. *Biometrics*, 50, 226–231. <https://doi.org/10.2307/2533213>
- Bell, S. (1999). *A beginner's guide to uncertainty of measurement*. National Physical Laboratory.
- Bhattacharya, C. B., & Sen, S. (2003). Consumer-company identification: A framework for understanding consumers' relationships with companies. *Journal of Marketing*, 67(2), 76–88. <https://doi.org/10.1509/jmkg.67.2.76.18609>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference* (2nd ed.). Springer-Verlag.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Camerer, C., & Yoon, C. (2015). Introduction to the Journal of Marketing Research special issue on neuroscience and marketing. *Journal of Marketing Research*, 52(4), 423–426. <https://doi.org/10.1509/0022-2437-52.4.423>
- Cho, G., Sarstedt, M., & Hwang, H. (2022). A comparative evaluation of factor- and component-based structural equation modelling approaches under (in)correct construct representations. *British Journal of Mathematical and Statistical Psychology*, 75(2), 220–251. <https://doi.org/10.1111/bmsp.12255>
- Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). LEA.
- Cronin, J. J., Brady, M. K., & Hult, G. T. M. (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing*, 76(2), 193–218. [https://doi.org/10.1016/S0022-4359\(00\)00028-2](https://doi.org/10.1016/S0022-4359(00)00028-2)
- Danilov, D., & Magnus, J. R. (2004). On the harm that ignoring pretesting can cause. *Journal of Econometrics*, 122(1), 27–46. <https://doi.org/10.1016/j.jeconom.2003.10.018>
- Danks, N. P., Sharma, P. N., & Sarstedt, M. (2020). Model selection uncertainty and multimodel inference in partial least squares structural equation modeling (PLS-SEM). *Journal of Business Research*, 113, 13–24. <https://doi.org/10.1016/j.jbusres.2020.03.019>
- de Jong, M. G., & Pieters, R. (2019). Assessing sensitive consumer behavior using the item count response technique. *Journal of Marketing Research*, 56(3), 345–360. <https://doi.org/10.1177/0022243718821312>
- Du, S., Bhattacharya, C. B., & Sen, S. (2007). Reaping relational rewards from corporate social responsibility: The role of competitive positioning. *International Journal of Research in Marketing*, 24(3), 224–241. <https://doi.org/10.1016/j.ijresmar.2007.01.001>
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–75. <https://doi.org/10.1214/ss/1177013815>
- Fisher, W. P., & Stenner, A. J. (2017). Towards an alignment of engineering and psychometric approaches to uncertainty in measurement: Consequences for the future. *18th International Congress of Metrology*, 12004, 1–9. <https://doi.org/10.1051/metrology/201712004>
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Sage.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science data-dependent analysis—A “garden of forking paths”—Explains why many statistically significant comparisons don't hold up. *American Scientist*, 102(6), 460. <https://doi.org/10.1511/2014.111.460>
- Geweke, J., & Meese, R. (1981). Estimating regression models of finite but unknown order. *International Economic Review*, 22(1), 55–70. [https://doi.org/10.1016/0304-4076\(81\)90091-9](https://doi.org/10.1016/0304-4076(81)90091-9)
- Glaveli, N. (2021). Two countries, two stories of CSR, customer trust and advocacy attitudes and behaviors? A study in the Greek and Bulgarian Telecommunication Sectors. *European Management Review*, 18(1), 151–166. <https://doi.org/10.1111/emre.12417>
- 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. <https://doi.org/10.1007/s11747-017-0517-x>
- Hair, J. F., Hult, T., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.
- Hayduk, L. A. (1994). Personal space: Understanding the simplex model. *Journal of Nonverbal Behavior*, 18(3), 245–260. <https://doi.org/10.1007/BF02170028>
- He, H., & Li, Y. (2011). CSR and service brand: The mediating effect of brand identification and moderating effect of service quality. *Journal of Business Ethics*, 100(4), 673–688. <https://doi.org/10.1007/s10551-010-0703-y>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hussey, I. (2021). A method to streamline p-hacking. *Meta-Psychology*, 5, 1–3. <https://doi.org/10.15626/MP.2020.2529>
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69(1), 81–99. <https://doi.org/10.1007/BF02295841>
- JCGM. (2012). *International vocabulary of metrology—Basic and general concepts and associated terms (VIM)* (3rd ed.). JCGM. https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, 23(2), 121–145. <https://doi.org/10.1111/j.2044-8317.1970.tb00439.x>
- Kennedy, P. (2008). *A guide to econometrics* (6th ed.). Blackwell.
- Lienggaard, B., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2021). Prediction: Coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling. *Decision Sciences*, 52(2), 362–392. <https://doi.org/10.1111/dec.12445>
- Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Physica.
- Lubke, G. H., & Campbell, I. (2016). Inference based on the best-fitting model can contribute to the replication crisis: Assessing model selection uncertainty using a bootstrap approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4), 479–490. <https://doi.org/10.1080/10705511.2016.1141355>
- Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, 13(2), 103–123. <https://doi.org/10.1002/job.4030130202>
- Öberseder, M., Schlegelmilch, B. B., Murphy, P. E., & Gruber, V. (2014). Consumers' perceptions of corporate social responsibility: Scale development and validation. *Journal of Business Ethics*, 124, 101–115. <https://doi.org/10.1007/s10551-013-1787-y>
- Oliveira, P. M., Guerreiro, J., & Rita, P. (2023). Neuroscience research in consumer behavior: A review and future research agenda. *International Journal of Consumer Studies*, 46(5), 2041–2067. <https://doi.org/10.1111/ijcs.12800>
- Pérez, A., & Del Bosque, I. R. (2015). Corporate social responsibility and customer loyalty: Exploring the role of identification, satisfaction and type of company. *Journal of Services Marketing*, 29(1), 15–25. <https://doi.org/10.1108/JSM-10-2013-0272>
- Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, 17(1), 1–14. <https://doi.org/10.1037/a0026804>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ray, S., Danks, N. P., & Calero Valdez, A. (2022). *SEMinR: Building and estimating structural equation models (version 2.3.2)*. <https://cran.r-project.org/web/packages/semnr/index.html>
- Rigdon, E. E., & Sarstedt, M. (2022). Accounting for uncertainty in the measurement of unobservable marketing phenomena. In H. Baumgartner & B. Weijters (Eds.), *Review of marketing research* (Vol. 19, pp. 53–73). Emerald.
- Rigdon, E. E., Sarstedt, M., & Becker, J. M. (2020). Quantify uncertainty in behavioral research. *Nature Human Behaviour*, 4(4), 329–331. <https://doi.org/10.1038/s41562-019-0806-0>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- 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. <https://doi.org/10.1016/j.jbusres.2016.06.007>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Sharma, P. N., Sarstedt, M., Shmueli, G., & Thiele, K. O. (2019). PLS-based model selection: The role of alternative explanations in IS research. *Journal of the Association for Information Systems*, 20(4), 346–397. <https://doi.org/10.17005/1.jais.00538>
- Sharma, P. N., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S. (2021). Prediction-oriented model selection in partial least squares path modeling. *Decision Sciences*, 52(3), 567–607. <https://doi.org/10.1111/dec.12329>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Sweeten, G. (2020). Standard errors in quantitative criminology: Taking stock and looking forward. *Journal of Quantitative Criminology*, 36, 263–272. <https://doi.org/10.1007/s10940-020-09463-9>
- Turker, D. (2009). Measuring corporate social responsibility: A scale development study. *Journal of Business Ethics*, 85(4), 411–427. <https://doi.org/10.1007/s10551-008-9780-6>
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. <https://doi.org/10.3758/BF03206482>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond $p < .05$. *The American Statistician*, 73(Suppl. 1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- Zeithalm, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60(2), 31–46. <https://doi.org/10.1177/002224299606000203>

How to cite this article: Rigdon, E., Sarstedt, M., & Moisescu, O.-I. (2023). Quantifying model selection uncertainty via bootstrapping and Akaike weights. *International Journal of Consumer Studies*, 47(4), 1596–1608. <https://doi.org/10.1111/ijcs.12906>

APPENDIX A: COVARIANCE MATRIX OF THE OBSERVED VARIABLES IN THE SIMULATION EXAMPLE

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
y1	0.96									
y2	0.51	0.97								
y3	0.30	0.23	1.25							
y4	0.26	0.20	0.78	1.15						
y5	0.25	0.11	0.44	0.39	1.25					
y6	0.21	0.12	0.41	0.30	0.77	1.16				
y7	0.04	-0.01	0.17	0.10	0.37	0.37	1.13			
y8	-0.02	-0.05	0.21	0.14	0.36	0.31	0.70	1.06		
y9	0.09	0.10	0.15	0.13	-0.01	0.03	0.30	0.24	1.24	
y10	0.17	0.13	0.24	0.20	0.02	0.01	0.22	0.17	0.75	1.12

APPENDIX B: COMPLETE R SCRIPT FOR THE SIMULATION EXAMPLE

```

# load lavaan package--install first, if necessary
library(lavaan)
# specify population model with parameter values
simmod<-'
#factor-factor
f2~0.5*f1
f3~0.5*f2
f4~0.5*f3
f5~0.4*f4+0.2*f1+0.2*f2+0.3*f3
#factor-indicator
f1=~0.8*y1+0.7*y2
f2=~0.8*y3+0.7*y4
f3=~0.8*y5+0.7*y6
f4=~0.8*y7+0.7*y8
f5=~0.8*y9+0.7*y10
y1~~0.36*y1
y2~~0.51*y2
y3~~0.36*y3
y4~~0.51*y4
y5~~0.36*y5
y6~~0.51*y6
y7~~0.36*y7
y8~~0.51*y8
y9~~0.36*y9
y10~~0.51*y10
'
# generate initial sample of 500 observations consistent with
simmod
simdat<-simulateData(model=simmod,std.lv=T,model.type =
"SEM",sample.nobs=500)

# specify candidate model 1, with path from f1 to f5
mod1<-'
#factor-factor
f2~f1
f3~f2
f4~f3
f5~f4 + f1 # but no direct paths from f2 or f3
#factor-indicator
f1=~y1+y2
f2=~y3+y4
f3=~y5+y6
f4=~y7+y8
f5=~y9+y10
'
# estimate candidate model 1
mod1.out<-sem(model=mod1,data=simdat)
# get AIC for model 1
# remember that R is case-sensitive almost always
mod1.AIC<-AIC(mod1.out)
# specify candidate model 2, with direct path from f2 to f5
mod2<-'
#factor-factor
f2~f1
f3~f2
f4~f3
f5~f4 + f2 # but no direct paths from f1 or f3
#factor-indicator
f1=~y1+y2
f2=~y3+y4
f3=~y5+y6
f4=~y7+y8
f5=~y9+y10
'
# estimate candidate model 2

```

```

mod2.out<-sem(model=mod2,data=simdat)
# get AIC value for candidate model 2
mod2.AIC <- AIC(mod2.out)
# specify candidate model 3, with direct path f3 to f5
mod3<- '
#factor-factor
f2~f1
f3~f2
f4~f3
f5~f4+f3 # but no direct paths from f1 or f2
#factor-indicator
f1=~y1+y2
f2=~y3+y4
f3=~y5+y6
f4=~y7+y8
f5=~y9+y10
'
# estimate candidate model 3
mod3.out<-sem(model=mod3,data=simdat)
# get AIC
mod3.AIC<-AIC(mod3.out)
# Find deltaAIC for the three candidate models
mod1.delta<-mod1.AIC-min(mod1.AIC,mod2.AIC,mod3.AIC)
mod2.delta<-mod2.AIC-min(mod1.AIC,mod2.AIC,mod3.AIC)
mod3.delta<-mod3.AIC-min(mod1.AIC,mod2.AIC,mod3.AIC)
# Calculate Akaike weights for the three candidate models
mod1.w<-exp(-.5*mod1.delta)/(exp(-.5*mod1.delta)+exp
(-.5*mod2.delta)+exp(-.5*mod3.delta))
mod2.w<-exp(-.5*mod2.delta)/(exp(-.5*mod1.delta)+exp
(-.5*mod2.delta)+exp(-.5*mod3.delta))
mod3.w<-exp(-.5*mod3.delta)/(exp(-.5*mod1.delta)+exp
(-.5*mod2.delta)+exp(-.5*mod3.delta))
# Set total number of bootstrap resamplings
totalboot<-10000
# Set number of resamplings for each candidate model,
# rounded to a whole number
mod1.n<-round(totalboot*mod1.w,0)
mod2.n<-round(totalboot*mod2.w,0)

```

```

mod3.n<-round(totalboot*mod3.w,0)
# Run the bootstrapping--this may take a while, with 10,000 total
resamplings
mod1.boot<-bootstrapLavaan(mod1.out,mod1.n)
mod2.boot<-bootstrapLavaan(mod2.out,mod2.n)
mod3.boot<-bootstrapLavaan(mod3.out,mod3.n)
# Extract the desired column of parameter estimates from the
output matrix
# Look at the matrix first to determine which column contains
# the desired parameter estimate
# It could be a different column in different output objects
param.boot1<-mod1.boot[,4]
param.boot2<-mod2.boot[,4]
param.boot3<-mod3.boot[,4]
## Check the number of replications obtained. If you have
had some
## nonconvergence, you may need to run some replacement
resamplings
library(psych)
describe (param.boot1)
describe (param.boot2)
describe (param.boot3)
# lavaan's bootstrap output is a matrix, so the extracted column
# is also a matrix. It will be convenient o have it be a data.frame,
# so convert it
param.boot1<-as.data.frame(param.boot1)
param.boot2<-as.data.frame(param.boot2)
param.boot3<-as.data.frame(param.boot3)
# it is also convenient to give the column of data a name
colnames(param.boot1)<- "f4f5"
colnames(param.boot2)<- "f4f5"
colnames(param.boot3)<- "f4f5"
# combine the parameter estimates into one variable
param.boots<-rbind(param.boot1,param.boot2,param.boot3)
# Get mean and standard deviation (standard error) of bootstrap
# parameter estimates across all resamplings
mean(param.boots$f4f5)
sd(param.boots$f4f5)

```


APPENDIX C: MEASUREMENT MODEL ASSESSMENT FOR THE EMPIRICAL EXAMPLE

Tables C1 and C2

TABLE C1 Reliability and convergent validity assessment.

Constructs and items		L	CA	CR	AVE
ENV	<i>Perceived brand environmental responsibility</i>				
ENV1	Tries to reduce its impact on the environment	0.796	0.884	0.886	0.661
ENV2	Strives to minimize the consumption of limited resources	0.878			
ENV3	Works diligently to use environmentally friendly materials	0.837			
ENV4	Is concerned with waste management and recycling	0.733			
CBI	<i>Customer-brand identification</i>				
CBI1	I am interested in what others think about this brand	0.672	0.839	0.846	0.649
CBI2	I feel angry when someone criticizes this brand	0.883			
CBI3	I feel good when someone praises this brand	0.846			
SAT	<i>Customer satisfaction</i>				
SAT1	I am very satisfied with this brand's products	0.758	0.867	0.869	0.689
SAT2	This brand's products are exactly what I need	0.872			
SAT3	My choice to buy this brand was a very good one	0.856			
LOY	<i>Customer loyalty</i>				
LOY1	This brand is my first choice, compared to other brands	0.767	0.828	0.841	0.639
LOY2	I will continue to be a customer of this brand	0.867			
LOY3	In the future I plan to purchase more products from this brand	0.759			

Note: AVE, average variance extracted; CA, Cronbach's alpha; CR, composite reliability; L, standardized factor loadings.

TABLE C2 Discriminant validity assessment.

	ENV	CBI	SAT	LOY
ENV	0.813	0.373	0.327	0.384
CBI	0.370	0.806	0.314	0.405
SAT	0.311	0.297	0.830	0.782
LOY	0.336	0.368	0.779	0.799

Note: Bold = square root of AVE; Regular = correlation; Italic = HTMT value (Henseler et al., 2015).

AUTHOR BIOGRAPHIES

Edward Rigdon is a Marketing Round Table Professor in the Robinson College of Business at Georgia State University in Atlanta, USA. He holds a PhD from the University of Alabama. His writing has appeared in premier journals across the fields of marketing, information systems and quantitative psychology. He co-founded *semnet*, an email list devoted to structural equation modelling, which was the focus of his dissertation and his early research. More recently, he has also worked toward the development of an uncertainty-centric research paradigm for the social sciences.

Marko Sarstedt is a professor of marketing at the Ludwig-Maximilians-University Munich (Germany) and an adjunct research

professor at Babeş-Bolyai-University Cluj-Napoca (Romania). His main research interest is the advancement of research methods to further the understanding of consumer behaviour. His research has been published in *Nature Human Behaviour*, *Journal of Marketing Research*, *Journal of the Academy of Marketing Science*, *Multivariate Behavioral Research*, *Organizational Research Methods*, *MIS Quarterly* and *Psychometrika*, among others. Marko has been repeatedly named member of Clarivate Analytics' Highly Cited Researchers List. In March 2022, he was awarded an honorary doctorate from Babeş-Bolyai-University Cluj-Napoca for his research achievements and contributions to international exchange.

Ovidiu-Ioan Moisescu is a professor of marketing and branding at the Marketing Department, Faculty of Economics and Business Administration at Babeş-Bolyai University of Cluj-Napoca, Romania. He completed his PhD in Marketing at the West University of Timișoara (Romania), as well as a postdoctoral research project at Babeş-Bolyai University of Cluj-Napoca (Romania) and Corvinus University of Budapest (Hungary). His research focuses on consumer behaviour, brand loyalty, CSR, and methodological issues in PLS-SEM.