

Should A Rule of Thumb be used to Calculate PLS-SEM Sample Size

Chanta Jhantasana

**Faculty of Management Science, Valaya Alongkorn Rajabhat University
under the Royal Patronage, Pathum Thani 13180, Thailand**

***Corresponding author: Chanta@vru.ac.th**

Received: February 18, 2022 Revised: March 30, 2023 Accepted: April 18, 2023

Abstract

Partial least square structural equation modeling (PLS-SEM) is more commonly used in marketing research because small sample sizes can be used. The main advantage is that the rule of thumb is often used to determine sample size, but the results may be underpowered. Therefore, appropriate sample size is still required to reach the acquired power of 0.80, which should be adequate to avoid false positive and false adverse effects arising from sample sizes that are too large or too small.

This research investigates the impact of different sample sizes on the power of analysis, the effect size, and the significance level of the model fitness and parameter estimation process. Many methods are used to generate study sample sizes, such as minimum R^2 , ten-time rule, inverse square root method, Marsh et al. method, Soper method, and Yamane method. The rule of thumb methods of minimum R^2 and ten-time rule generate sample sizes that are too small and inappropriate for PLS-SEM.

However, the findings have shown that PLS-SEM can be effective with small sample sizes, but the sample size should be more significant than that generated by the rule-of-thumb methods. The appropriate sample size for this study was 50, with a power of 0.81 and an effect size (f^2) ranging between 0.437 and 0.506.

Keywords: Sample size, Effect size, Statistical power, Partial least square structural equation model, Rule of thumb

Introduction

Social and behavioral sciences are increasingly using structural equation modeling (Bollen, et al, 2022). It allows researchers to evaluate ideas and hypotheses (Sarstedt et al., 2021). The simulation of structural equations has two methods. These are covariance-based structural equation modeling (CB-SEM: Jöreskog, 1993) and partial least square structural equation modeling (PLS-SEM: Hair & Alamer, 2022). Although CB-SEM is more commonly used for marketing research, PLS-SEM has recently become more popular (Magno et al.,

2022). The PLS-SEM provides the significant benefit of using small sample data to validate models and refine parameter estimates, mainly when the sample size is smaller than the indicators used in the measurement model. Wold (1989) showed that using 10 samples with 27 indicators was sufficient to fit the structural model. The rule of thumb is currently more hastily applied to sample size analysis, particularly the ten-times and minimum R^2 rule (Hair et al., 2017). However, some researchers (e.g., Kock & Hadaya, 2018; Rönkkö & Evermann, 2013) have criticized because the ten-times rule may result in small sample sizes misspecifying the measurement model. Though PLS-SEM can be used with small samples, it should be good for its population to be represented. Inadequate sample sizes will likely create inaccuracies, particularly for heterogeneous groups (Hair et al., 2017). A small sample size study would result in low statistical power and a small size of the effect (e.g., Aguirre-Urreta & Rönkkö, 2015; Kock & Hadaya, 2018). In sample size estimation, type-I and type-II errors should be considered. A false positive or type-I error (alpha error) is likely to be generated by too large a sample size leading to the belief that the null hypothesis is incorrect when this is true. A false negative or type-II error is usually caused by too small a sample size (beta error). It happens when the null hypothesis is incorrectly accepted, thus creating insufficient statistical power. Therefore, an appropriate sample size must be used to ensure expected effect sizes and a minimum acceptable power of 0.80 (Cohen, 1988).

As a rule of thumb, size and statistical power evaluations are usually problematic (Aguirre-Urreta & Rönkkö, 2015). A critical factor in determining the statistical power to conduct an a priori PLS-SEM study is the sample size (Aguirre-Urreta & Rönkkö, 2015; Cohen, 1988; Kock & Hadaya, 2018). Inappropriate sample size can result in bias and distort findings of path coefficients. Higher statistical power means that PLS-SEM is more likely to classify relationships as relevant when present in the population (Sarstedt & Mooi 2014). However, due to the lack of software support, few statistical power reports from PLS-SEM are available (Aguirre-Urreta & Rönkkö, 2015; Baroudi & Orlikowski, 1989; Chin, 1998; Marcoulides et al., 2009; Ringle et al., 2012). Most estimates of sufficient statistical power are based on non-validated rules of thumb (Aguirre-Urreta & Rönkkö, 2015). This analysis will evaluate the appropriate sample size, which is the smallest sample size affecting the model to be generated by at least 0.8 of statistical power (Kock, 2017), effect size of 0.15 (Cohen, 1988) and a path coefficient of 0.20 (Hair et al., 2010). Therefore, this study will examine which methodology is best suited to producing the appropriate sample size for the intention of e-repurchase. The research is an experiment involving eight methods of sample size 10 (Wold, 1982), 30 (ten-times method), 38 (minimum R^2), 160 (Inverse-square root), 100 (Soper, 2021; Reinartz et al., 2009), 50, and 200 (March et al., 1998), and 400 (Yamane, 1980), respectively. Ultimately, this study will also demonstrate how WrapPLS can gain statistical power.

Literature review

Sample size estimate

Several rules of thumb exist for estimating sample sizes in the structural equation modeling literature. For example, Bentler and Chou (1987); Bollen (1989) used 5-10 times the

number of path coefficients and ten times the number of indicators (Nunnally & Bernstein, 1994) to decide their necessary sample sizes. Reinartz et al. (2009) used a sample size of 100, with sufficient statistical power. Hair et al. (2017) used ten times the number of indicators considered to have a substantial effect size and sufficient statistical power. However, these rules can give rise to sample sizes that need to be bigger. Therefore, the appropriate sample size calculation should also consider its related statistical power and effect size, as the sample size is a function of alpha beta, and effect size (Cunningham & McCrum-Gardner, 2007).

Statistical power is the probability $(1 - \beta)$ that the null hypothesis will be rejected if it is false to avoid type II errors. Regarding statistical power, the significance level is the minimum probability that the population is correlated with the sample. The statistical power should not be less than 0.80, which means that beta (β) or type II errors should not exceed 0.20 (Cohen, 1988). The statistical power will decrease if beta (beta) increases, so the sample size needs to increase. The statistical power is equal to the sample size of the effect size $\lambda = f^2 N$ (, where λ is the statistical power, f^2 is the effect size, and N is the sample size (Ellis, 2010). Therefore, if an increase in sample size increases statistical power, small sample sizes will produce a lower path coefficient.

On the other hand, an appropriate sample size produces a higher path coefficient (Chin & Newsted, 1999). However, the sample size should be prominent if the data is not a normal distribution (Marcoulides & Saunders, 2006). Ensure adequate statistical power, the significance of the path coefficient, and acceptable effect size, the sample size should also be appropriate and manageable. When a researcher seeks justification for choosing an ad hoc interest in his inquiry, the ideal power value (0.80) can be ignored (Cohen, 1988).

The PLS-SEM effect size are also used in this analysis more commonly used by f^2 (Cohen 1988), estimated based on $R^2 / (1 - R^2)$. For example, effect sizes of 0.02, 0.15, and 0.35 are defined as small, medium, and large (Cohen, 1992). PLS-SEM may use the smallest sample size, which still produces a statistical power of 0.80 if the path coefficient and effect size is large enough (Hair et al., 2017). On the other hand, this condition fails if the effect size is below medium. The (appropriate) sample size is then related to statistical power (0.80), effect size (0.15), and path coefficient (0.20). The present research explores differences in sample size using PLS-SEM to determine the appropriate sample size, mainly the using the rule of thumb, including the sample size of Wold (1982). Larger sample sizes, determined by several variables, particularly the effect size, power, and alpha level functional to the sample size, are also used in this analysis (Cunningham & McCrum-Gardner, 2007).

Table1 Recommended sample sizes for PLS-SEM at a statistical power of 0.80

Maximum Number of Arrows Pointing at a Construct	Significance Level											
	1%				5%				10%			
	Minimum R ²				Minimum R ²				Minimum R ²			
	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75
2	158	75	47	38	110	52	33	26	88	41	26	21

Maximum Number of Arrows Pointing at a Construct	Significance Level											
	1%				5%				10%			
	Minimum R ²				Minimum R ²				Minimum R ²			
	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75	0.10	0.25	0.50	0.75
3	176	84	53	42	124	59	38	30	100	48	30	25
4	191	91	58	46	137	65	42	33	111	53	34	27
5	205	98	62	50	147	70	45	36	120	58	37	30
6	217	103	66	53	157	75	48	39	128	62	40	32
7	228	109	69	56	166	80	51	41	136	66	42	35
8	238	114	73	59	174	84	54	44	143	69	45	37
9	247	119	76	62	181	88	57	46	150	73	47	39
10	256	123	79	64	189	91	59	48	156	76	49	41

Source: Cohen (1992); Hair et al. (2017)

Minimum R²

Hair et al. (2017) proposes an alternative method that yields a substantially larger sample than the ten-times rule. This method is based on the Cohen (1992) power tables for the least square regression in which the table lists the minimum required sample sizes based on three elements, namely the maximum number of arrows pointing to the latent variable, the significance level used, and the minimum R² in the model. Table 1 emphasis on 0.05 levels of significance and 0.80 statistical power. The maximum number of arrows pointing to a latent variable in this study is three, with a sample size of 38.

Ten-times rule

The proposed rule of thumb by Barclay, Higgins, and Thompson (1995) is that the minimum sample size for a study should be ten times the number of indicators in the most complex formative model, or ten times the largest number of indicators for independent variables in the structural model, depending on which one is bigger. The ten-times rule does not consider the path coefficient's size. Ringle et al. (2018) discovered that the sample sizes vary in many fields, for example, human resources, 142.5 are also used in this analysis service business about 332; marketing, about 211.3 strategic management, about 154.9, and supply chain management about 274.4. Rigdon (2016) found that using a small sample size did not achieve adequate statistical power in the results. Goodhue et al. (2006) found that experiments using the ten-times rule to measure sample size produce a small effect size and affect the statistical power leading to errors.

Inverse square root

Kock & Hadaya (2018) suggested an inverse square root method. The formula of this method is $N \succ \left(\frac{2.48}{|bm|} \right)^2$, where $|bm|$ is the absolute path coefficient. This method is based on the Monte Carlo estimate, which assumes that p is equal to or less than 0.05, that the effect size is equal to or greater than 0.15, that the statistical power is at least 0.80, and that their multicollinearity. Warp PLS 6.0 will measure the sample size using the square inverse root method.

Marsh et al. (1998) method

Marsh et al. (1998) calculated the sample size based on proper solution and non-convergence. Non-convergence means programming a computer that cannot evaluate a pattern within 250 iterations. Common factors are included in the PLS path models, and so-called Heywood cases are possible (Krijnen, Dijkstra & Gill, 1998), meaning that the model suggested one or more negative variances. Too small a sample may cause Heywood cases to occur or not carry the typical factor structure for a specific set of indicators (Henseler, Hubona, & Ray, 2015), requiring the correct solution (Kolenikov & Bollen, 2012).

Table 2 Proper solution and non-convergence

Indicators per Latent	Type	Sample Size				
		50	100	200	400	1,000
2	Proper Solution	32.8	55.6	55.6	82.4	93.0
	Non-Convergence	33.1	12.6	12.6	2.4	-
3	Proper Solution	85.4	97.8	97.8	100	99.0
	Non-Convergence	0.9	-	-	-	-
4	Proper Solution	99.1	99.6	99.6	100	100
	Non-Convergence	-	-	-	-	-
6	Proper Solution	100	100	100	100	100
	Non-Convergence	-	-	-	-	-
12	Proper Solution	100	100	100	100	100
	Non-Convergence	-	-	-	-	-

Source: Piriyaikul (2019)

In Table 2, Piriyaikul (2019) explains the method of Marsh et al. The and the indicator for the latent variable is essential elements. Indicators per latent are 2, 3, 4, 6, and 12, while sample sizes are 50, 100, 200, 400, and 1,000. For an indicator per latent of two and a sample size of 50, a proper solution of 32.8 (more is better) and a non-convergence of 33.1 (less is better) will be obtained, which is insufficient. Increasing the sample size to 400 reveals that the appropriate non-convergence solution is 82.4 and 2.4, respectively. Indicators per latent equal to or more than four can use sample sizes from 50 to 200. If the number of indicators per latent is between 3 and 12, the acceptable sample size is between 100 and 200. On the other hand, a sample size from 400 to 1,000 is too large since the proper solution and non-convergence are already stable at a sample size of 200.

Soper's (2021) method

This method measures sample size and can be found at <https://www.danielsoper.com/>. The measurement factor consists of the effect, the statistical power, the number of indicators and latent variables, and the degree of significance. This research uses an effect size of 0.30, a

statistical power of 0.80, fifteen indicators, three latent variables, and a significance value of 0.05, resulting in a minimum sample size for the model structure of 100.

Yamane's table

The most used tool for measuring sample size in Thailand is Yamane (1980). The smallest sample size in the Yamane table is 400, which yields an error of than 5% when the population is infinite.

E-Recovery, e-trust & e-repurchase intention literature

E-Recovery service quality

Online communication is more technologically based and can result in quality service failures, such as site bugs, failure to receive items, failure to access all pages, and loss of passwords. E-recovery service quality is often seen as a passive approach to improving customer satisfaction (Zeithaml et al., 2006). Suppose the firm offers a fast response to customers it will need for confidence, commitment, loyalty, and word-of-mouth advantages. Parasuraman et al. (2005) proposed the quality of e-recovery services, which assess a firm's responsiveness, compensation, and contact. Responsiveness entails that consumers can return goods and services quickly. Compensation entails a process to pay via the website, and contact means that consumers can contact the organization's representatives in various ways, by telephone and online.

E-trust

E-trust is the customer's perception of the website used to send products and services according to their expectations, including trust in information technology and the website (Bart et al., 2005). E-trust relies heavily on information technology and its efficacy. That means that E-trust relies more on the website's functionality and design. E-trust will weaken, producing negative word of mouth, if the Internet infrastructure is unstable. E-trust is central to repurchasing customers (Chiu et al., 2010). Lee and Turban (2001) found that consumers will only buy when they have confidence in the website.

E-repurchasing intention

The intention of an e-repurchase is defined as a client's intention to purchase the same product or service online on more than one occasion (Ibzan et al., 2016). The profitability and competitiveness of companies can be strengthened by repurchasing. E-purchase stems from the concept of traditional marketing. The online market, however, has the advantage of easier repurchasing. Some elements influence repurchasing customers, such as trust, comfort, satisfaction, and risk (Dash & Saji, 2007). Consumers who are satisfied with a firm and have positive word of mouth comments about product and service quality often end up doing repurchases (Rust & Zahorik, 1993). Claudia (2012) found that satisfaction allows consumers to make repeat purchases. However, one risk that significantly affects repurchasing is the price of the products or services. Customers may risk buying cheaper products rather than more ~~than~~ expensive ones (Chiu et al., 2014).

Hypothesis and Conception Framework

According to the literature review, this study uses indicators of responsiveness, compensation, and contact to assess the quality of e-recovery services and to study the path analysis of e-recovery service quality and e-repurchase intention, as shown in the conceptual framework in Figure1. There are three hypotheses.

H₁: The e-recovery service quality is positively and direct to e-trust.

H₂: The e-recovery service quality is positively and direct to e-repurchase intention.

H₃: The e-trust is positive direct to e-repurchase intention.

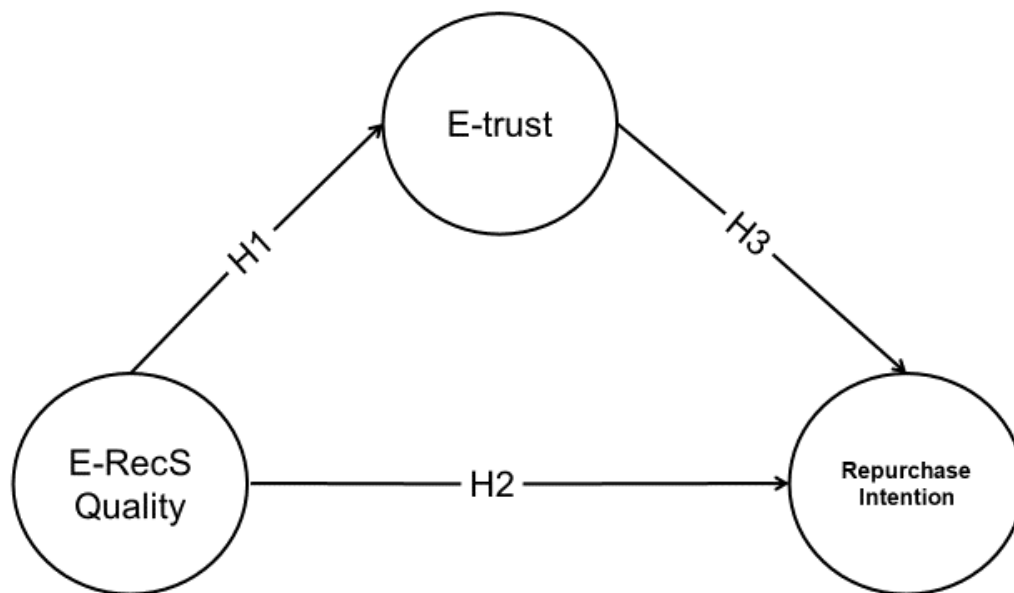


Figure1 Hypothesis and conceptual framework

Methodology

Population and sample

This research investigates the analysis of e-recovery service quality and e-trust on repurchase intentions among students at Valayalongkorn Rajabhat University. A sample of 400 students was collected using an accidental sampling method. Smaller samples of sizes 10, 30, 38, 50, 100, 160, and 200 were randomly chosen from an original sample of 400 individuals, which was selected using systematic random sampling. While these smaller samples are likely representative of the population, the precision of estimates based on them may be lower, resulting in a higher chance of sampling error. Appropriate statistical methods should be used to assess the validity of any conclusions drawn from smaller samples. The sample sizes selected were 10 (Wold, 1982), 30 (ten-times rule), 38 (minimum R^2), 160 (inverse square root), 100 (Soper, 2021), 50, 10, and 200 (Marsh et al., 1998) and 400 (Yamane, 1980).

Measurement

Respondents were asked to complete an e-recovery service quality service questionnaire (Parasuraman et al., 2005), an e-trust questionnaire (Valvi & West, 2013; Safa & Von Solms, 2016; Chu & Yuan, 2013), and an e-repurchase intention questionnaire (Pee et al., 2018; Khalifa & Liu, 2007). E-recovery service quality consisted of three sub-constructs variables: compensation, responsibility, and contract. The questionnaire utilized a 5-point Likert scale, ranging from Strongly Disagree. The authors translated the questionnaire into the Thai language. Following this, English and Thai experts, who had never seen the questionnaire translated it back into English. Two reviewers prepared and checked the final language edition comparing original version and the newly translated one.

Table 3 Indicators

Indicators
e-Recovery Service Quality
COMP1: This site compensates me for the problems it creates.
COMP2: It compensates me if I do not get what I ordered on time.
COMP3: It picks up the items that I want to return from home or business.
RESP1: This site handles returns well.
RESP2: It addresses problems promptly.
RESP3: The company solves problems quickly in my opinion.
CONT1: This site provides the company's telephone number.
CONT2: This site has online customer service representatives.
CONT3: This provides the opportunity to talk to a live person if a question arises.
e-Trust
TRUS1: I can trust that this website works well.
TRUS2 :In terms of my privacy, I trust this website.
TRUS3: I trust the information provided on this website.
e-Repurchase intention
REPU1 :If I repurchase the product, I'd like to purchase it from the same website.
REPU2 :I want to reuse this website for my next purchase.
REPU3: I want to revisit the website to purchase products soon.

PLS-SEM

PLS-SEM is one of the structural equation model methods based on the variance family. PLS-SEM consists of a collection of regression systems that can be evaluated by constructing component weights and pathways between exogenous and endogenous structures. The algorithm is built in several steps, which iteratively calculate the latent variable scores in the first stage. The second stage resolves the measurement model, while the third stage defines the structural model parameters (Hair et al., 2017). Hair et al. (2018) summarizes the various

reasons for using PLS-SEM. These include predictability, complex models, exploratory theory, formative structures, data artifacts, secondary data, small sample sizes, and a lack of normal distribution. PLS-SEM's primary rationale is based on non-normal data, small samples, and formative, calculated constructs (Hair et al., 2014). This paper focuses primarily on the small size of the sample.

This study focused on the path analysis of the quality of e-recovery services and e-trust for e-repurchase purposes using PLS-SEM WrapPLS 6 (Kock, 2019). PLS-SEM consists of measurement and structural models. The measurement model studies the relationship between the indicators and the latent variables. In contrast, the structural model studies the relationship between the latent variables and a recursive model of endogenous and exogenous variables. The algorithms estimate all latent variables by loading each latent variable to determine the path coefficient and the parameters.

The reflective measurement model's accuracy is considered valid and consistent. Validity in this analysis applies to convergent and discriminatory validity. Consistent validity finds loadings greater than 0.708 (Henseler et al., 2015). Discriminant Validity explores the association between the latent variable below AVE's square root (Fornell & Larcker, 1981). Reliability considers Cronbach's Alpha and composite reliability, which can reach 0.70 (Ali et al., 2018).

The structural model's quality depends on the path coefficient, the R^2 , and the size of the effect (f^2). The path coefficients are expected to be greater than 0.20. (Chin, 1998). Three levels represent R^2 (0.25, 0.50, and 0.75) as weak, neutral, and sufficient, respectively (Hair et al., 2010). The is dependent on those three levels. An effect size of 0.02 means a common effect and a small sample size. An effect of 0.15 refers to a neutral impact and medium sample size. If the effect size is 0.35, this means a significant impact and a large sample size (Cohen, 1988).

For WrapPLS 6.0, the overall model's quality is considered using the path coefficient, R^2 , block VIF, and full collinearity VIF. If the VIF is below 5, it implies non-collinearity, and averages below 3.3 indicate the best results (Kock, 2019). The Goodness of Fit Index (GoF) is an explanatory power measurement of the model proposed by Tenenhaus et al. (2005). The product's square root is between the average Group Index and the R-square measure. Wetzels et al. (2009) suggested the following thresholds for GoF: small if 0.1; medium if 0.25, or more significant if 0.30 or more.

Results

The goodness of fit

The results show that the GoF for all sample sizes is high. The path coefficients for all sample sizes range from 0.487 to 0.560 and are statistically significant at 0.001. R^2 s range from 0.490 to 0.660 and are statistically significant at 0.001. The averages VIF and full collinearity VIF of all sample sizes is below 3. The model fitness for all sample sizes is between 0.639 and 0.750. The results shown in Table 4.

Table 4 The goodness of fit

n	Average Path Coefficient	Average R ²	Average Block VIF	Average Full Collinearity VIF	Tenenhaus GoF (GoF)
10	0.550***	0.640***	2.377	3.509	0.740
30	0.524***	0.579***	2.137	2.478	0.723
38	0.560***	0.660***	2.116	3.454	0.750
50	0.540***	0.606***	1.811	3.198	0.710
100	0.498***	0.503***	1.506	2.425	0.655
160	0.490***	0.493***	1.532	2.226	0.639
200	0.500***	0.521***	2.111	2.251	0.670
400	0.487***	0.490***	1.703	2.182	0.647

Remark: *p<0.10, **p<0.05, and ***p<0.001

Table 5 Parameters of the measurement model for all sample sizes

Indicators	Loading of Each Sample Size							
	10	30	38	50	100	160	200	400
Resp1	0.901	0.852	0.831	0.814	0.847	0.827	0.811	0.834
Resp2	0.901	0.854	0.845	0.845	0.825	0.775	0.826	0.836
Resp3	0.894	0.862	0.854	0.825	0.832	0.863	0.819	0.821
Comp1	0.584	0.871	0.793	0.791	0.838	0.843	0.852	0.848
Comp2	0.623	0.871	0.797	0.791	0.831	0.805	0.837	0.836
Comp3	0.605	0.846	0.774	0.737	0.808	0.861	0.830	0.803
Cont1	0.819	0.903	0.806	0.797	0.884	0.871	0.875	0.869
Cont2	0.806	0.902	0.814	0.815	0.875	0.856	0.870	0.864
Cont3	0.883	0.889	0.838	0.820	0.858	0.827	0.878	0.855
Trust1	0.967	0.985	0.951	0.929	0.943	0.887	0.952	0.943
Trust2	0.967	0.993	0.981	0.972	0.972	0.957	0.979	0.974
Trust2	0.993	0.984	0.956	0.952	0.951	0.948	0.968	0.967
Repu1	0.988	0.980	0.972	0.949	0.950	0.940	0.946	0.946
Repu2	0.997	0.990	0.991	0.985	0.980	0.973	0.982	0.981
Repu3	0.997	0.991	0.982	0.975	0.959	0.961	0.974	0.972

Remark: *All loadings are significant at the 0.001, level.

Measurement model

Appropriate loading reliability is more than 0.708. The results show that most loads are more significant than 0.708 (see Table 6), excluding those with a sample size of 10 for which loads of compensation are below the criterion. The internal consistency of Cronbach's Alpha is considered. Composite reliability refers to the reliability of the total latent variables being equal to or greater than 0.70. Internal consistency and composite reliability are more significant than 0.90 (Table 6). It is a "shared variance indicator between the observed variables used as a latent construct indicator" (Fornell & Larcker, 1981). Convergent validity is based on an average extracted variance (AVE) equal to or greater than 0.50. The results show that the AVEs of all sample sizes are more significant than 0.50, which means that all models demonstrate convergent validity (Table 6). Discriminant validity is the validity of the measuring instrument in the latent variable. The results show that the correlation between the latent variables of all the sample sizes is lower than in the, except for the sample size of 10. As a result, almost all models gain discriminative validity (Table 7).

Table 6 Internal consistency reliability

n	Latent	Full Collinearity VIFs	Construct Reliability		Convergent Validity
			Cronbach's Alpha (α)	Composite Reliability	AVE
10	Recovery	0.293	0.921	0.936	0.625
	Trust	3.072	0.975	0.983	0.952
	Repurchase	4.522	0.994	0.996	0.989
30	Recovery	2.278	0.961	0.966	0.761
	Trust	2.755	0.987	0.991	0.975
	Repurchase	2.401	0.987	0.991	0.975
38	Recovery	2.691	0.938	0.948	0.668
	Trust	3.412	0.960	0.974	0.926
	Repurchase	4.258	0.981	0.988	0.964
50	Recovery	2.370	0.932	0.943	0.647
	Trust	3.110	0.947	0.966	0.905
	Repurchase	4.115	0.968	0.979	0.940
100	Recovery	1.881	0.950	0.957	0.713
	Trust	2.410	0.952	0.969	0.913
	Repurchase	2.985	0.961	0.974	0.927
160	Recovery	1.709	0.939	0.950	0.703
	Trust	2.438	0.923	0.951	0.867
	Repurchase	2.529	0.955	0.971	0.918

n	Latent	Full Collinearity VIFs	Construct Reliability		Convergent Validity AVE
			Cronbach's Alpha (α)	Composite Reliability	
200	Recovery	2.411	0.950	0.957	0.714
	Trust	2.298	0.965	0.977	0.934
	Repurchase	2.043	0.966	0.978	0.936
400	Recovery	1.848	0.948	0.956	0.707
	Trust	2.312	0.959	0.973	0.924
	Repurchase	2.386	0.964	0.977	0.933

Table 7 Discriminant validity

Latent Variables		Recovery	Trust	Repurchase
10	Recovery	0.790		
	Trust	0.701**	0.976	
	Repurchase	0.809***	0.819***	0.994
30	Recovery	0.872		
	Trust	0.721***	0.987	
	Repurchase	0.670***	0.737***	0.987
38	Recovery	0.817		
	Trust	0.720***	0.962	
	Repurchase	0.784***	0.834***	0.982
50	Recovery	0.805		
	Trust	0.660***	0.951	
	Repurchase	0.757***	0.822***	0.970
100	Recovery	0.845		
	Trust	0.575***	0.955	
	Repurchase	0.678***	0.761***	0.963
160	Recovery	0.838		
	Trust	0.592***	0.931	
	Repurchase	0.611***	0.749***	0.958
200	Recovery	0.845		
	Trust	0.715***	0.966	
	Repurchase	0.671***	0.651***	0.967
400	Recovery	0.841		
	Trust	0.621***	0.961	
	Repurchase	0.636***	0.724***	0.966

Remark: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.001$

Structural model

The structural model's quality depends on the coefficient of determination (R^2), the Stone-Geisser predictive relevance test (Q^2), the effect size, and the path coefficient. Table 8 shows that the R^2 for Trust and Repurchase is greater than the medium level, except for the sample sizes of 10, 38, and 50 in repurchase. All trust and purchase Q^2 s are far from zero significantly and have the maximum values for the sample sizes of 10, 38, and 50 in the repurchase. The R^2 and Q^2 values indicate the same story throughout the model.

Table 8 displays the study of skewness and kurtosis and the statistical power for all sample sizes. It shows that the values for skewness and kurtosis range from 0.008 to -0.493. A dataset has a normal distribution if its skewness falls between -2 and +2 and its kurtosis falls between -7 and +7 (Byrne, 2010; Hair et al., 2010). The statistical power values are more significant than 0.80 when the sample size is 50 or higher.

Table 8 Parameters to measure structural model

n	Latent variables	R^2	Q^2	Skewness	Kurtosis	Power
10	Recovery	-	-	0.448	-1.064	<0.60 ⁺
	Trust	0.510	0.547	0.451	-1.740	
	Repurchase	0.770	0.787	-0.419	-1.233	
30	Recovery	-	-	-0.045	-0.876	<0.50
	Trust	0.544	0.534	-0.627	-0.409	
	Repurchase	0.615	0.619	-0.613	-0.513	
38	Recovery	-	-	0.008	-0.300	0.716
	Trust	0.535	0.537	0.250	-0.733	
	Repurchase	0.785	0.784	0.142	-0.954	
50	Recovery	-	-	0.175	-0.108	0.810
	Trust	0.437	0.445	-0.210	-0.368	
	Repurchase	0.775	0.774	-0.020	-0.589	
100	Recovery	-	-	-0.093	-0.762	0.974
	Trust	0.332	0.343	-0.324	-0.371	
	Repurchase	0.675	0.681	-0.091	-0.548	
160	Recovery	-	-	0.112	-0.977	0.961
	Trust	0.355	0.357	-0.436	0.516	
	Repurchase	0.631	0.629	-0.219	-0.277	
200	Recovery	-	-	0.049	-0.966	>0.99 ⁺⁺
	Trust	0.517	0.516	-0.493	-0.374	
	Repurchase	0.524	0.518	-0.361	-0.499	
400	Recovery	-	-	0.049	-0.879	>0.99 ⁺⁺
	Trust	0.389	0.390	-0.446	-0.288	
	Repurchase	0.590	0.587	-0.277	-0.525	

Note: + The value may be less than 0.60, but it cannot be calculated because the samples are less than the indicators.++ The highest value is 0.99.

Table 9 shows the effect size (f^2), greater than 0.15, or medium for all direct, indirect, and total effect sizes. Thus, all sample sizes' three path coefficients are interpreted as large effect sizes.

Table 9 The effect size (f^2) of direct and indirect path

n	Path	f^2 of direct path	f^2 of indirect path	Total
10	H ₁ : Recovery-Trust	0.510		0.510
	H ₂ : Recovery-Repurchase	0.351	0.293	0.644
	H ₃ : Trust-Repurchase	0.419		0.419
30	H ₁ : Recovery-Trust	0.544		0.544
	H ₂ : Recovery-Repurchase	0.186	0.284	0.470
	H ₃ : Trust-Repurchase	0.429		0.429
38	H ₁ : Recovery-Trust	0.535		0.535
	H ₂ : Recovery-Repurchase	0.287	0.338	0.625
	H ₃ : Trust-Repurchase	0.499		0.499
50	H ₁ : Recovery-Trust	0.437		0.437
	H ₂ : Recovery-Repurchase	0.269	0.302	0.571
	H ₃ : Trust-Repurchase	0.506		0.506
100	H ₁ : Recovery-Trust	0.332		0.332
	H ₂ : Recovery-Repurchase	0.248	0.220	0.468
	H ₃ : Trust-Repurchase	0.428		0.428
160	H ₁ : Recovery-Trust	0.355		0.355
	H ₂ : Recovery-Repurchase	0.171	0.225	0.396
	H ₃ : Trust-Repurchase	0.495		0.495
200	H ₁ : Recovery-Trust	0.517		0.517
	H ₂ : Recovery-Repurchase	0.261	0.189	0.450
	H ₃ : Trust-Repurchase	0.264		0.264
400	H ₁ : Recovery-Trust	0.389		0.389
	H ₂ : Recovery-Repurchase	0.179	0.221	0.410
	H ₃ : Trust-Repurchase	0.411		0.411

Table 10 shows that each path's coefficients or direct effects are strong or more significant than 0.20. All the path coefficients are significant. This implies that the three hypotheses are accepted for all sample sizes.

The study suggests considering the mediating role of e-trust between e-recovery service quality and e-repurchase intention. The three paths demonstrate a significant and positive

relationship, indicating that e-trust partially mediates e-recovery service quality and e-repurchase intention (Nitzl et al., 2016).

Table 10 Hypothesis test of each sample size

n	Path	Direct Effect	Indirect Effect	Total Effect	Assumption
10	H ₁ : Recovery-Trust	0.714***		0.714***	Accept
	H ₂ : Recovery-Repurchase	0.431**	0.360**	0.791***	Accept
	H ₃ : Trust-Repurchase	0.504***		0.504***	Accept
30	H ₁ : Recovery-Trust	0.737***		0.737***	Accept
	H ₂ : Recovery-Repurchase	0.272**	0.415***	0.687***	Accept
	H ₃ : Trust-Repurchase	0.563***		0.563***	Accept
38	H ₁ : Recovery-Trust	0.732***		0.732***	Accept
	H ₂ : Recovery-Repurchase	0.363***	0.429***	0.792***	Accept
	H ₃ : Trust-Repurchase	0.587***		0.587***	Accept
50	H ₁ : Recovery-Trust	0.661***		0.661***	Accept
	H ₂ : Recovery-Repurchase	0.355***	0.398***	0.753***	Accept
	H ₃ : Trust-Repurchase	0.602***		0.602***	Accept
100	H ₁ : Recovery-Trust	0.576***		0.576***	Accept
	H ₂ : Recovery-Repurchase	0.362***	0.321***	0.683***	Accept
	H ₃ : Trust-Repurchase	0.557***		0.577***	Accept
160	H ₁ : Recovery-Trust	0.596***		0.596***	Accept
	H ₂ : Recovery-Repurchase	0.273***	0.359***	0.632***	Accept
	H ₃ : Trust-Repurchase	0.602***		0.602***	Accept
200	H ₁ : Recovery-Trust	0.719***		0.719***	Accept
	H ₂ : Recovery-Repurchase	0.388***	0.282***	0.670***	Accept
	H ₃ : Trust-Repurchase	0.392***		0.392***	Accept
400	H ₁ : Recovery-Trust	0.624***		0.624***	Accept
	H ₂ : Recovery-Repurchase	0.281***	0.348***	0.629***	Accept
	H ₃ : Trust-Repurchase	0.557***		0.557***	Accept

Remark: * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.001$

Discussion

In this study, the appropriate sample size for the PLS-SEM is the smallest sample capable of generating at least 0.80 of statistical power and an effect size of 0.15. This is not a bias in the parameters of the corresponding sample. The results show that the three loading (Compt1, Compt2, and Compt3) are below the criterion for sample size 10. However, the hypothesis is likely to accept all other parameters used. Prejudice may also have existed. The results for sample size 10 would be approved for general use of PLS-SEM with a loading cut-off from the model if it were below 0.708 (Henseler et al., 2015), while sample sizes 30 and 38 would be accepted as is. According to Cohen (1988), the statistical power for the three models with sample sizes of 10, 30, and 38 is below 0.80, which implies that the study has a high risk of producing false negative results (Ellis, 2010). Therefore, the appropriate and required sample size is 50, which generates a statistical power of 0.81. If we do not consider statistical power, the appropriate sample size starts at 30 because the relevant parameters and the effect size are quite good.

In this analysis, a sample size of 30 and 38 are made by the rule of thumb or by the ten-times and minimum R^2 methods. The model fitness indices obtained from the sample sizes are above the criterion. However, the statistical power is less than 0.80, which indicates that the results provide no confidence in the assumptions. As a result, the ten-times and minimum R^2 approaches produce inappropriate sample sizes. The appropriate sample size for this analysis begins at 50, which provides a statistical power of 0.81, with an effect size (f^2) between 0.437 and 0.571. The power of a sample size of 200 is around 0.99, while a sample size of 400 overpowers the study trend. Therefore, this analysis's acceptable sample size is between 50 and 200.

The study suggests that e-trust partially mediates in influencing e-repurchase intention between e-recovery service quality. This indicates that when e-trust is established, it can increase the likelihood of customers' e-repurchase intention. In other words, more than the quality of e-recovery services is required to ensure repeat purchases. However, e-trust can be a significant factor in building customers' confidence and trust in online recovery services, which may increase their willingness to repurchase from the same service provider.

Conclusion and recommendation

Practical implications

The results show that the rule of thumb produces too small a sample size, which means that the statistical power is less than 0.80; thus, we can avoid using this rule. This research indicates that the Kock and Hadaya (2018); Marsh et al. (1998); Soper (2021) methods are sufficient for the analysis of sample sizes. Kock and Hadaya (2018) offer an inverse square root method that can be used with WarpPLS to evaluate if it has the same application for statistical power testing. The methods of Marsh et al. (1998) play a significant role in Table 2,

in which the sample size depends on the number of indicators per latent variable. However, the calculation is less related to model parameters such as statistical power and effect size. However, Soper (2021) can be the simplest method for calculating sample size that can be accessed via website interaction. The consumer can fill out six factors that will generate the appropriate sample size in addition to the effect size, the statistical power, the number of latent variables, the number of indicators, and the beta level.

The study recommends that the smallest sample size be produced as a digit number on the WrapPLS and Daniel Soper websites. WrapPLS is a commercial program costing approximately 100 USD per year with a three-month free trial version. WrapPLS 7 has enhanced its various applications, particularly using Heterotrait-Monotrait (HTMT) to investigate discriminatory validity. WrapPLS is the only PLS-SEM software used to investigate statistical power. The website Danielsoper.com is provided by Dr. Daniel Soper of California State University (Fullerton), Professor of Information Technology and Decision Science. This website contains 106 free statistical calculators for sample size, statistical power, effect size, normal distribution, cumulative distribution function, Poisson distribution, etc. The next section discusses how to investigate power and the method of Inverse square Root using WarpPLS.

WrapPLS 6.0 is a useful program for investigating the statistical power of various sample sizes using the “Explore” menu, then the “Explore statistical power and the minimum sample size requirement” submenu. There are two parts. The first part of which consists of three boxes. The first box is the minimum absolute path coefficient of the model (ranging from 0.01 to 0.99). The second box is the significance used (ranging from 0.001 to 0.50). The third box is the power level requirements (ranging from 0.50 to 0.99). The second part consists of: the inverse square root method and the gamma-exponential method. The figures show the line graphs of the relation between the vertical statistical power and the horizontal sample size. The default values for each form’s three boxes and sample sizes are 0.197, 0.05, 0.80, 160, and 140, respectively.

The calculation begins using the smallest significance path coefficient of the PLS-SEM performance, such as 0.36, and then fills in the first box. The sample sizes for the inverse square root and gamma-exponential are reduced from 160 and 140 to 48 and 35, respectively. However, we concentrate primarily on the method of the inverse square root. The consequence is that 48 is less than the sample size of 50, so we must increase the statistical power. The default statistical power is 0.80. If we set it to 0.85, as derived from a sample size of 56, then the statistical power will decrease from 0.85 to 0.81, and the sample size will be 50, which is equal to the actual sample size.

Theoretical implications

The leading theory of this analysis is the small sample size (Hair et al., 2011), which is the outstanding character of PLS-SEM as suggested by most researchers (e.g., Hair et al., 2018). As an inventor, (Wold, 1982), Wold’s critical reason for constructing the PLS-SEM was small samples. Since PLS-SEM retains its core functionality, it does not simultaneously

evaluate all model parameters (Rigdon et al., 2017). Tenenhaus et al. (2004) argue that only partial model structures are estimated simultaneously. This study confirms that the relevant sample size stratum of 50 samples has been established.

The second theory concerns the statistical power used in this analysis to investigate the sample's size. Sample sizes depend on three factors: alpha, beta, and effect size (Cunningham & McCrum-Gardner, 2007). The research project and statistical test results are calculated using four related components: sample size, effect size, alpha level, and power. The sample size is the number of research units-accessible (e.g., people). The effect size refers to the relationship between the treatment (i.e., noise measurement) and the outcome variable (i.e., treatment salience). The power, on the other hand, refers to the likelihood of detecting a true effect of the treatment on the outcome variable, given the effect size and other factors such as the sample size and alpha level. Together, these concepts help to evaluate the strength of the relationship between the treatment and the outcome variable and the likelihood of detecting this relationship. The fourth value can be calculated for all three elements, such as the appropriate sample size for the analysis. If the effect size, alpha level, and power size could be precisely estimated, the sample size would be measured. This research confirms that statistical power is a key PLS-SEM sample size analysis indicator. Consequently, statistical power reporting must be the product of all PLS-SEM models; otherwise, the probability of error or model type I or II loses all meaning.

The last theory is the methodology theory used to measure the sample size. The sample size rule of thumb produces an unacceptable number of samples, particularly the ten-times rule. Some authors, such as Hair et al. (2017); Barclay et al. (1995), have constructed it differently, but the effect size is small. Also, the indicator used to calculate the rule of thumb for sample size is less related to the model parameters since the sample size, the statistical power, the effect size, and the alpha level are all related and dependent. An efficient method for investigating the required sample size must contain at least these four factors. WrapPLS 6 (Kock, 2019); Soper (2021) consist of these four parameters. In addition, Soper added the number of latent variables and the number of all the model indicators. This study confirms that the rule of thumb is not adequate to investigate sample size and must provide statistical power, effect size, and an alpha level method for investigating sample size. Inline, Morey and Lakens (2016) state that it is preferable to choose a sample size for the test that has sufficient power to assess the research interest's effect.

Limitation and future research

In this study, the quality of e-recovery services is considered a higher-order construct (HOC), consisting of responsibility, compensation, and contract as its first-order construct. In PLS-SEM, HOCs can be built using four methods: repeated indicator, two-stage, hybrid, and three-stage approaches. The first three approaches showed a high correlation between the first and second-order latent variables, especially in small sample sizes, which is one of the limitations of this study. However, a third-stage approach was suggested for HOC constructions, which can also assess model fitness, a critical weakness of the other three

methods (van Riel et al., 2017). The other downside is the need for more analysis for too-large sample sizes or the possibility of false positive or type-I errors in this test. Based on this analysis findings, sample sizes exceeding 500 can be an example of a false positive due to approximate overpower. The literature supports these issues, as stated by Kock and Hadaya (2018); Kock (2018).

Conclusion

In conclusion, our study highlights the importance of the impact of the use of a small sample size in PLS-SEM. Small sample sizes used in model fitness and parameter estimations can cause the results to contain errors. We found that the rule of thumb can produce wrong sample sizes, particularly the ten-time and minimum R^2 methods, while other methods can generate acceptable sample sizes. In future research, we propose investigating appropriate sample sizes for high-order construction analysis used in the three-stage approach. Since it is necessary to develop a sample size that guarantees sufficient statistical power, the Soper (2021) method will be more effective than the rule of thumb in determining the sample size.

When the statistical power of a study is lower than 0.80, the study has a high risk of producing false negative results, which can occur when the sample size is too small, the effect size is too small, or there is too much variability in the data. A low statistical power increases the chances of committing a Type II error, which is the failure to reject a false null hypothesis. This may result in a missed opportunity to uncover significant impacts and draw reliable conclusions from the study. Therefore, and conducting studies with adequate statistical power is recommended to ensure reliable and valid results.

References

- Aguirre-Urreta, M., & Rönkkö, M. (2015). Sample size determination and statistical power analysis in PLS using R: An annotated tutorial. *Communications of the Association for Information Systems*, 36(3), 33-51.
- Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., and 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.
- Baroudi, J. J., & Orlikowski, W. J. (1989). The problem of statistical power in MIS research. *MIS Quarterly*, 13(1), 87-106.
- Bart, Y., Shankar, V., Sultan, F., and Urban, G. L. (2005). Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing*, 69(4), 133-152.
- Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78-117.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modelling: A Multidisciplinary Journal*, 7(3), 461-483.
-

- Chin, W. W. (1998). *The partial least squares approach to structural equation modeling* (pp. 295-336). In Marcoulides, G. A. (Ed.). *Modern methods for business research*. Mahwah, NJ: Lawrence Erlbaum Associate.
- Chin, W. W., & Newsted, P. R. (1999). *Structural equation modeling analysis with small samples using partial least squares* (pp. 307-342). In Hoyle, R. H. (Ed.). *Statistical strategies for small sample research*. Thousand Oaks, CA: Sage Publications.
- Chiu, C., Huang, H., & Yen, C. (2010). Antecedents of trust in online auctions. *Electronic Commerce Research and Applications*, 9(2), 148-159.
- Chiu, C., Wang, E. T., Fang, Y., & Huang, H. (2012). Understanding customers' repeat purchase intentions in B2C e-commerce: The roles of utilitarian value, hedonic value, and perceived risk. *Information Systems Journal*, 24(1), 85-114.
- Chu, K.M., & Yuan, J. (2013). The effects of perceived interactivity on e-trust and e-consumer behaviors: The application of fuzzy linguistic scale. *Journal of Electronic Commerce Research*, 14(1), 124-136.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd eds.). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cunningham, J. B., & Mc-Crum Gardner, E. (2007). Power, effect, and sample size using GPower: practical issues for researchers and members of research ethics committees. *Evidence-Based Midwifery*, 5(4), 132-136.
- Dash, S., & Saji, K. B. (2008). The role of consumer self-efficacy and website social-presence in customers' adoption of B2C online shopping. *Journal of International Consumer Marketing*, 20(2), 33-48.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. London: UK: Cambridge University press.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350), 320-328.
- Goodhue, D., Lewis, W., & Thompson, R. (2006). *PLS, small sample size and statistical power in MIS research* (pp.202b-202b). In Proceedings of the 39th Hawaii international conference on system sciences. Los Alamitos, CA: IEEE Computer Society Press.
- Hair, F. J., Sarstedt, M., Hopkins, L., & G. Kuppelwieser, V. (2014). Partial least squares structural equation modeling (PLS-SEM). *European Business Review*, 26(2), 106-121.
- Hair, F. J., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2017). *Advanced Issues in partial least squares structural equation modeling*. Thousand Oaks, CA: SAGE Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. (7th eds.). Englewood Cliffs: Prentice-Hall.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed, a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139-152.
-

- Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced issues in partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: Sage.
- Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), 100027.
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, 116(1), 2-20.
- 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.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20, 277-319.
- Ibzan, E., Balarabe, F., & Jakada, B. (2016). Consumer satisfaction and repurchase intentions. *Developing Country Studies*, 6(2), 96-100.
- Jöreskog, K. G. (1993). *Testing structural equation models* (pp. 294-316). In Bollen, K. A., & Long, J. S. (Eds.). *Testing structural equation models*. Newbury Park: Sage.
- Khalifa, M., & Liu, V. (2007). Online consumer retention: contingent effects of online shopping habits and online shopping experience. *European Journal of Information Systems*, 16(6), 780-792.
- Kock, N. (2018). *Minimum sample size estimation in PLS-SEM: An application in tourism and hospitality research* (pp.1-16). In Ali, F., Rasoolimanesh, S. M., & Cobanoglu, C. (Eds.). *Applying partial least squares in tourism and hospitality research*. Bradford, UK: Emerald Publishing Limited.
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227-261.
- Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances. *Sociological Methods & Research*, 41(1), 124-167.
- Krijnen, W. P., Dijkstra, T. K., & Gill, R. D. (1998). Conditions for factor (in) determinacy in factor analysis. *Psychometrika*, 63(4), 359-367.
- Lee, M. K., & Turban, E. (2001). A trust model for consumer internet shopping. *International Journal of Electronic Commerce*, 6(1), 75-91.
- Magno, F., Cassia, F., & Ringle, C. M. (2022). A brief review of partial least squares structural equation modeling (PLS-SEM) use in quality management studies. *The TQM Journal*, 1754-2731.
- Marcoulides & Saunders. (2006). Editor's comments: PLS: A silver bullet? *MIS Quarterly*, 30(2), iii-ix.
- Marcoulides, Chin, & Saunders. (2009). A critical look at partial least squares Modelling. *MIS Quarterly*, 33(1), 171-175.
-

- Marsh, H. W., Hau, K., Balla, J. R., and Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181-220.
- Maxham, J. G. (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of Business Research*, 54(1), 11-24.
- Morey, R., & Lakens, D. (2016). *Why most of psychology is statistically unfalsifiable*. Retrieved from doi:10.5281/zenodo.838685
- Nitzl, C., Roldan, J. L., & Cepeda, G. (2016). Mediation analysis in partial least squares path modeling. *Industrial Management & Data Systems*, 116(9), 1849-1864.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd eds.). New York, NY: McGraw-Hill, Inc.
- Parasuraman, A., Zeithaml, V. A. & Malhotra, A. (2005). E-S-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research*, 7(10), 1-21.
- Pee, L., Jiang, J., & Klein, G. (2018). Signaling effect of website usability on repurchase intention. *International Journal of Information Management*, 39, 228-241.
- Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *SSRN Electronic Journal*, 26(4), 332-344.
- Rigdon, E. E. (1998). Advanced structural equation modeling: Issues and Techniques. *Applied Psychological Measurement*, 22(1), 85-87.
- Rigdon, E. E. (2016). Choosing PLS path modeling as an analytical method in European management research: A realist perspective. *European Management Journal*, 34(6), 598-605.
- 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*, 39(3), 4-16.
- Ringle, C. M., Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2018). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, 31(12), 1617-1643.
- Ringle, Sarstedt, & Straub. (2012). Editor's comments: A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, 36(1), iii-xiv.
- Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, 16(3), 425-448.
- Rust, R. T., & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of Retailing*, 69(2), 193-215.
- Safa, N. S., & Von Solms, R. (2016). An information security knowledge sharing model in organizations. *Computers in Human Behavior*, 57, 442-451.
- Sarstedt, M., & Mooi, E. (2014). *A concise guide to market research: The process, data, and methods using IBM SPSS Statistics*. Basingstoke, CA: Springer.
- Soper, D. S. (2021). A-priori sample size calculator for structural equation models [Software]. Retrieved from <https://www.danielsoper.com/statcalc/calculator.aspx?id=89>
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111-133.
-

- Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). A global goodness-of-fit index for PLS structural equation modeling. *Proceedings of the XLII SIS scientific meeting*, 739-742.
- Tompson, R., Barclay, D. W., & Higgins, C. A. (1995). The partial least squares approach to causal modeling: Personal computer adoption and uses as an illustration. *Technology Studies: Special Issue on Research Methodology*, 2(2), 284-324.
- Valvi, A. C., & West, D. C. (2013). E-Loyalty is not all about trust, the price also matters: extending expectation-confirmation theory in bookselling websites. *Journal of Electronic Commerce Research*, 14(1), 99-123.
- Van Riel, A. C., Henseler, J., Kemény, I., & Sasovova, Z. (2017). Estimating hierarchical constructs using consistent partial least squares. *Industrial Management & Data Systems*, 117(3), 459-477.
- Wold, H. (1982). *Soft modeling: The basic design and some extensions* (pp. 1-54). In Jöreskog, K. G., & Wold, H. (Eds.). *Systems under indirect observations, part II*. Amsterdam: North-Holland.
- Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2006). *Services marketing: Integrating customer focus across the firm* (4th eds.). Singapore: McGraw-Hill.