When Six Respondents Make the Model: Extreme-Response Clusters and the Fragility of PLS-SEM Results in Tourism Technology Adoption Research

Haslinda Musa, Nur Zafirah A Khadar, Nurulizwa Abdul Rashid, Samer Ali Hussein Al-Shami, Fadhlur Rahim Azmi

Faculty of Technology Management and Technopreneurship, Universiti Teknikal Malaysia Melaka, , Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.1017PSY0073

Received: 21 October 2026; Accepted: 26 October 2026; Published: 16 November 2026

ABSTRACT

Tourism technology adoption studies rely heavily on Likert-scale surveys analyzed with partial least squares structural equation modelling (PLS-SEM). Such data are often strongly skewed towards agreement. This study shows how a very small cluster of uniformly negative respondents can manufacture the reliability, validity, structural, predictive and configurational results that researchers report, and how the standard diagnostics fail to detect it. Survey data from 304 Malaysian respondents on the adoption of a solar-powered tourist boat (the ePenambang, Kuala Terengganu) were analyzed with an extended Technology Acceptance Model. Every analysis was run with and without six respondents who disagreed with all 35 items. A dose-response analysis, bootstrap composition analysis, permutation null models and a Monte Carlo simulation (2,000 replications) were used to separate substantive signal from the influence of the extreme cluster. With all 304 cases, the model met every conventional criterion: $\alpha = 0.81–0.91$, all ten paths significant and $R^2 = 0.85$ for intention. Without the six cases (2% of the sample), α fell to between −0.60 and 0.61, nine of ten paths became non-significant and R^2 fell to 0.24. Not one of 1,000 bootstrap resamples excluded all six cases. Necessary condition effect sizes fell from 0.46–0.72 to 0–0.11, a multigroup difference by education vanished, and fsQCA solutions were reproduced by a randomly permuted outcome. In simulation, adding six all-disagree cases to pure noise produced $R^2 = 0.83$, $\alpha = 0.84$ and HTMT above 0.90 for every construct pair. Standard careless-response indices did not flag the six cases; a multivariate outlier screen did. Researchers, reviewers and editors in tourism should require sensitivity analyses that report results with and without uniformly extreme respondents, report within-cluster correlations, and treat very high R^2 and reliability in ceiling-affected data as a warning rather than a strength. A ten-point reporting checklist is provided. To our knowledge, this is the first study to document, with real tourism data and simulation, how a small extreme-response cluster propagates through the full modern PLS-SEM toolkit, including NCA, fsQCA, MICOM/MGA, IPMA and PLSpredict.

Keywords: PLS-SEM; influential observations; careless responding; ceiling effects; necessary condition analysis; fsQCA; robustness; sustainable tourism technology

INTRODUCTION

Partial least squares structural equation modelling (PLS-SEM) has become one of the most widely used analytical methods in tourism and hospitality research (do Valle & Assaker, 2016; Ali et al., 2018). Its appeal is practical: it handles non-normal data, complex models and modest samples, and it now comes with a rich toolkit of extensions such as necessary condition analysis (NCA; Richter et al., 2020), combined PLS-SEM and fuzzy-set qualitative comparative analysis (fsQCA; Rasoolimanesh et al., 2021), measurement invariance and multigroup analysis (Henseler et al., 2016), importance–performance map analysis (IPMA; Ringle & Sarstedt, 2016) and out-of-sample prediction (Shmueli et al., 2019). Guidelines on how to report these analyses are widely followed (Hair et al., 2019; Becker et al., 2023).

Much less attention has been paid to the data that enter these models. Technology adoption surveys, including many conducted in Malaysia (Musa et al., 2016; Mustafa et al., 2017; Azmi et al., 2020; Safri & Musa, 2022; Musa et al., 2025), typically ask respondents to rate positively worded statements about a new, socially desirable technology on short Likert scales. Such data are prone to acquiescence and ceiling effects (Baumgartner & Steenkamp, 2001), so most respondents cluster at the top of the scale. When almost everyone agrees with almost everything, the little covariance that exists can be dominated by a handful of respondents who answer very differently, whether they are genuine rejecters of the technology or careless responders (Meade & Craig, 2012; Ward & Meade, 2023).

Influential observations are a well-recognised threat in regression analysis (Aguinis et al., 2013), and robustness checks for PLS-SEM have been proposed (Sarstedt et al., 2020). However, no study has yet shown how a small cluster of extreme respondents can propagate through the entire modern PLS-SEM workflow, from reliability and discriminant validity to NCA, fsQCA and multigroup analysis, or whether the diagnostics that researchers routinely apply can detect this propagation. Because these statistics share a common origin in the item correlation matrix, the same cluster that biases one statistic can bias all of them simultaneously.

This study addresses this gap with a worked case and an accompanying simulation. It uses a tourism technology adoption dataset (N = 304) in which six respondents (2%) disagree with every statement. Three research questions are addressed:

1. RQ1: To what extent can the measurement, structural, predictive and configurational evidence found in a typical PLS-SEM adoption study be attributed to a small number of uniformly extreme respondents?

2. RQ2: Do commonly used diagnostics, such as bootstrap confidence intervals, the HTMT ratio, common method bias tests, MICOM and careless-responding indices, reveal such dependence?

3. RQ3: Does the pattern generalise beyond this dataset, and what should researchers disclose to guard against it?

The study makes three contributions. First, it documents the fragility of every major PLS-SEM output to 2% of the sample, using real tourism data. Second, it shows by simulation that such a cluster can create a textbook-quality model from pure noise. Third, it offers a practical reporting checklist for tourism researchers, reviewers and editors.

Background and Propositions

PLS-SEM in Tourism Technology Adoption Research

Tourism researchers have adopted PLS-SEM rapidly, often citing its suitability for non-normal data and predictive aims (do Valle & Assaker, 2016; Ali et al., 2018). Technology adoption studies, typically built on the Technology Acceptance Model (Davis, 1989) and its extensions, are among its most common applications. A recent stream of guidance encourages researchers to go beyond path coefficients: to assess necessity with NCA (Richter et al., 2020), configurations with fsQCA (Pappas & Woodside, 2021; Rasoolimanesh et al., 2021), group heterogeneity with MICOM and multigroup analysis (Henseler et al., 2016), managerial priorities with IPMA (Ringle & Sarstedt, 2016) and predictive power with PLSpredict (Shmueli et al., 2019). Critics have warned that PLS can capitalize on chance and produce inflated estimates (Rönkkö et al., 2016), but these warnings have focused on the estimator rather than on the data.

Ceiling Effects, Response Styles and Careless Responding

Respondents differ in how they use rating scales. Acquiescent and extreme response styles vary across individuals and cultures and can distort means, correlations and structural relationships (Baumgartner & Steenkamp, 2001). Surveys about desirable, pro-environmental technologies invite agreement, which pushes responses against the top of the scale. Within such a ceiling, true differences between respondents are compressed and item correlations shrink.

At the same time, some respondents give little attention to the survey. Careless or insufficient-effort responding is usually detected with indices such as long-string analysis (the longest run of identical answers), intra-individual response variability and Mahalanobis distance (Meade & Craig, 2012; Curran, 2016; Ward & Meade, 2023). In a positively worded survey, a careless respondent who disagrees with everything and a genuine rejecter look the same, and both stand far from the agreeing majority.

The distinction matters for how such respondents should be handled. If they are careless responders, their answers carry no information about the constructs and should not shape the estimates. If they are genuine rejecters of the technology, they represent a real, if small, segment of the population whose views deserve to be retained. Because the two cannot be told apart from the response pattern alone, the appropriate default is not to delete such cases but to report results with and without them (Aguinis et al., 2013). This study follows that principle throughout.

Influential Observations

An influential observation is one whose omission changes substantive conclusions (Aguinis et al., 2013). In correlation-based methods, a few cases lying far from the main cluster can create a strong correlation between two variables that is absent within the cluster, simply because the outlying cases are low on both. Because PLS-SEM derives every statistic from the item correlation matrix, such cases can affect reliability, validity and path coefficients alike. Bootstrapping offers no safeguard: when resampling with replacement, the probability that a given resample contains none of the k influential cases is $(1 - k/N)^N$, which for k = 6 and N = 304 is about 0.2% (Efron & Tibshirani, 1993).

Propositions

These arguments lead to six propositions, summarized in Table 1.

Table 1: Propositions

Proposition	Statement
P1. Measurement	A small extreme-response cluster inflates reliability (α, CR, AVE) and inter-construct correlations, so that discriminant validity fails while convergent validity appears excellent.
P2. Method bias	Common method bias diagnostics (Harman, full-collinearity VIF) are driven by the same cluster and cannot distinguish it from method variance.
P3. Bootstrapping	Bootstrap confidence intervals do not reveal the dependence, because almost every resample contains the cluster.
P4. Structural model	Path coefficients, indirect effects and R^2 depend on the cluster in a dose-response manner.
P5. Advanced extensions	NCA, fsQCA, MICOM/MGA, IPMA and PLSpredict inherit the dependence, so their results do not provide independent corroboration.
P6. Detection	Univariate careless-response indices do not detect the cluster, whereas a multivariate outlier screen and case-removal sensitivity analysis do.

METHOD

Data

The case dataset comes from an online survey (Google Forms, October–December 2024) of 304 Malaysian travelers and tourism stakeholders about a solar-powered river boat, the ePenambang, in Kuala Terengganu. Respondents rated 35 positively worded statements on a five-point scale (1 = strongly disagree, 5 = strongly agree), measuring seven constructs with five items each: socioeconomic status (SS), environmental awareness (EA), technological proficiency (TP), perceived usefulness (PU), perceived ease of use (PEOU), subjective norms (SN) and adoption intention (AI). There were no missing values. The same data have been used in substantive studies (Musa et al., 2025), this study treats them only as a methodological case.

The data are typical of adoption surveys about desirable technologies: 76.2% of all answers were 5 and 97.2% were 4 or 5. Six respondents (2.0%; respondent IDs 3, 5, 26, 28, 29 and 30 in the supplementary file) answered 1 or 2 on every item, with a mean of 1.32, against a mean of 4.76 for the other 298. Five of the six were women, four were under 30, and all six were in the two lowest education codes.

Focal Model

The focal model is the extended TAM specification common in this literature: SS, EA and TP predict PU and PEOU; PEOU predicts PU; and PU, PEOU and SN predict AI. All constructs are measured reflectively (Mode A). The model was estimated with the PLS algorithm using the path weighting scheme, in a Python implementation that reproduces SmartPLS defaults.

Analytic Design

Five complementary lenses were used to separate signal from the influence of the six cases:

1. Case-removal comparison. Every analysis was run on the full sample (N = 304) and without the six cases (n = 298), following outlier-handling recommendations to report results both ways (Aguinis et al., 2013).

2. Dose-response. The model was re-estimated after removing every possible subset of the six cases (64 models), to show how each statistic changes as the cluster shrinks.

3. Bootstrap composition. For 1,000 bootstrap resamples, the number of draws of the six cases was recorded and related to the resample's R^2.

4. Permutation nulls. The intention items were randomly permuted 200 times to obtain the R^2 expected by chance. For fsQCA, the outcome was permuted 200 times to test whether solutions differed from those obtained for a random outcome.

5. Monte Carlo simulation. A third population tested the reverse case: the noise items were reflected ($6 - x$) to create a negatively skewed dataset, to which 0, 3, 6, 9 or 15 synthetic respondents answering uniformly 4 or 5 were added from the 298 non-extreme respondents: the real responses, and pure noise in which every item was shuffled independently to destroy all relationships while keeping the response distributions. For each, 200 bootstrap samples were drawn and 0, 3, 6, 9 or 15 synthetic respondents with random answers of 1 or 2 were added (2,000 replications). R^2, mean α and the share of HTMT ratios above 0.90 were recorded.

Extensions were assessed with their standard procedures: NCA with the CE-FDH ceiling and 1,000 permutations (Dul, 2016; Richter et al., 2020); fsQCA with three calibration schemes and standard consistency thresholds (Pappas & Woodside, 2021); MICOM and permutation MGA by education level (Henseler et al., 2016); IPMA (Ringle & Sarstedt, 2016); and PLSpredict with 10 folds (Shmueli et al., 2019). Careless-response screening used long-string analysis, intra-individual response variability and Mahalanobis distance (Curran, 2016).

RESULTS

Table 2: Summary of Key Results with and Without the Six Extreme Respondents

Statistic	**Full sample (N = 304)**	**Without the six cases (n = 298)**
Cronbach's α	0.81–0.91	−0.60 to 0.61
Composite reliability	0.87–0.93	0.00–0.22
HTMT ratios above 0.90	15 of 21	1 of 21 (11 undefined)
Harman's first factor	58.9% of variance	8.4% of variance

Full-collinearity VIF	2.25–13.93	1.07–2.04
Significant structural paths	10 of 10	1 of 10 (SN → AI)
R^2 (adoption intention)	0.85	0.236
Q^2predict	0.49–0.61	−0.02 to 0.20
NCA effect sizes (d)	0.46–0.72	0.00–0.11
MGA by education	2 significant differences	No differences
IPMA performance	87.6–94.5	Undefined

The Geometry of the Data

Figure 1 shows the core problem. The 298 agreeing respondents form a tight cluster in the upper-right corner, within which the correlation between the mean of the six predictors and intention is 0.02. The six extreme respondents sit alone in the lower-left corner. With them included, the same correlation is 0.91. Mean inter-item correlations within each scale tell the same story: 0.47–0.67 at N = 304, but −0.07 to 0.24 without the six cases.

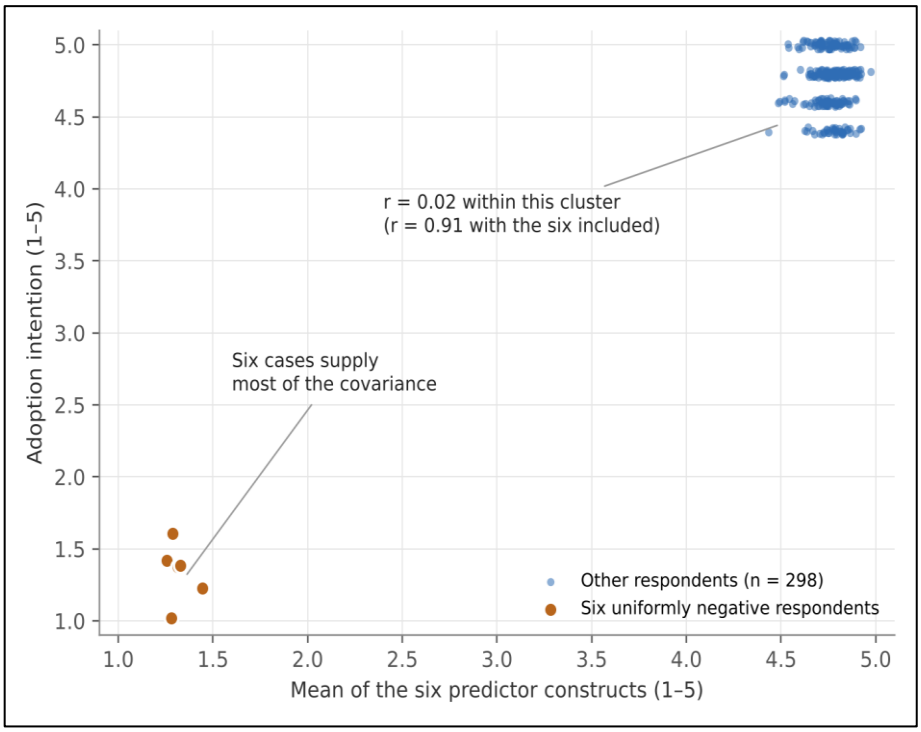


Figure 1: Predictor Means Against Adoption Intention: Two Clusters, One Correlation

Measurement Model (P1)

With all cases, the measurement model looked excellent (Table 3). All α values exceeded 0.80, CR exceeded 0.87 and AVE exceeded 0.57. Discriminant validity failed, with 15 of 21 HTMT ratios above 0.90 and seven above 1 (Henseler et al., 2015). Without the six cases, every scale lost its internal consistency: α ranged from −0.60 to 0.61, CR from 0.00 to 0.22 and AVE from 0.18 to 0.36, and HTMT could not be computed for 11 of 21 pairs because within-scale correlations were negative. P1 is supported: the cluster created both the convergent validity and the discriminant validity problem.

Table 3: Measurement Quality with and Without the Six Extreme Respondents

Construct	**α**	**CR**	**AVE**	**α**	**CR**	**AVE**
	(N = 304)	**(N = 304)**	**(N = 304)**	**(n = 298)**	**(n = 298)**	**(n = 298)**

SS	0.809	0.872	0.576	0.611	0.038	0.214
EA	0.879	0.922	0.704	−0.191	0.218	0.263
TP	0.866	0.905	0.655	0.368	0.047	0.176
PU	0.906	0.932	0.734	0.073	0.09	0.274
PEOU	0.9	0.93	0.729	0.116	0.002	0.299
SN	0.888	0.922	0.703	−0.596	0.121	0.202
AI	0.866	0.904	0.654	0.093	0.219	0.357
HTMT > 0.90 (of 21)	15			1(11 undefined)		

Note: SS = socioeconomic status; EA = environmental awareness; TP = technological proficiency; PU = perceived usefulness; PEOU = perceived ease of use; SN = subjective norms; AI = adoption intention.

Common Method Bias Diagnostics (P2)

At N = 304, a single factor (Harman's test) explained 58.9% of item variance, and Kock's (2015) full-collinearity VIFs ranged from 2.25 (SS) to 13.93 (EA). A researcher following conventional rules (Podsakoff et al., 2003; Kock, 2015) would conclude that common method bias is present. Without the six cases, the first factor explained only 8.4% of variance and all VIFs fell to 1.07–2.04, below every threshold. The diagnostics were therefore detecting the cluster, not a method factor shared by all respondents. P2 is supported.

Structural Model (P4)

Figures 2 and 3 show the estimated PLS path models. At N = 304, all ten paths were significant, R^2 reached 0.902 for PU, 0.866 for PEOU and 0.850 for AI, and both indirect effects of environmental awareness on intention were significant. Without the six cases, only SN → AI remained significant (β = 0.460, p < 0.001), and R^2 for AI fell to 0.236 (Table 4). This remaining R^2 exceeded the permutation-null 95th percentile (0.068), so a modest social-norm signal is genuine. Note that the bootstrap confidence interval for R^2 excludes zero in both samples, as it must for a statistic that cannot be negative; it is not evidence of explanatory power.

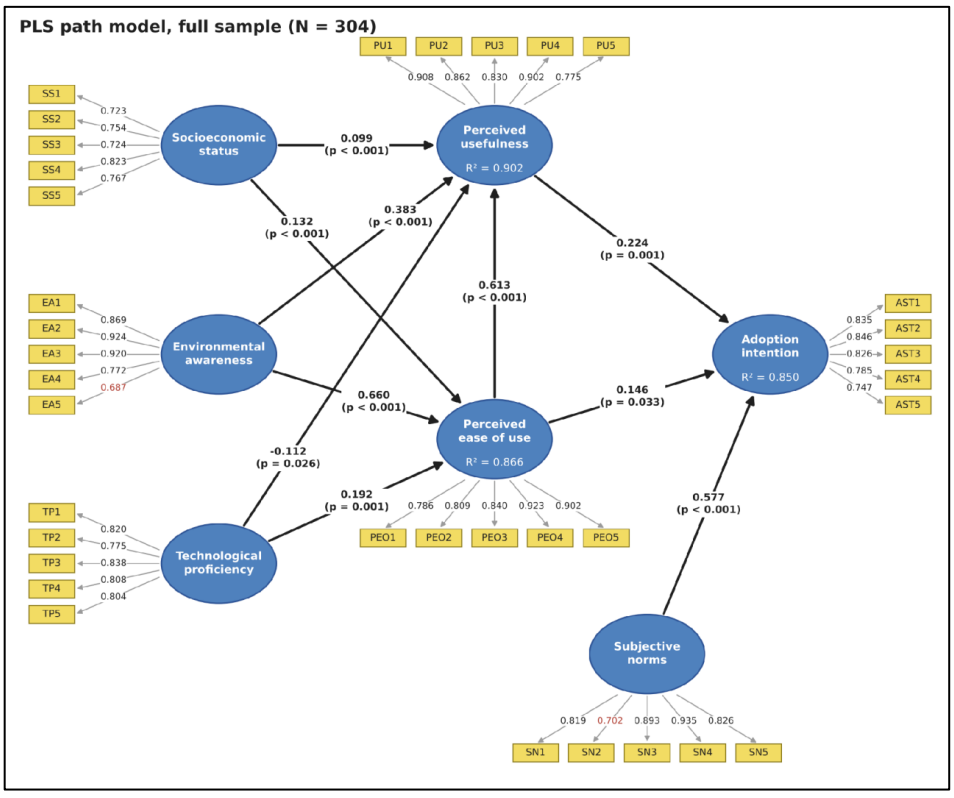


Figure 2: PLS Path Model, Full Sample (N = 304)

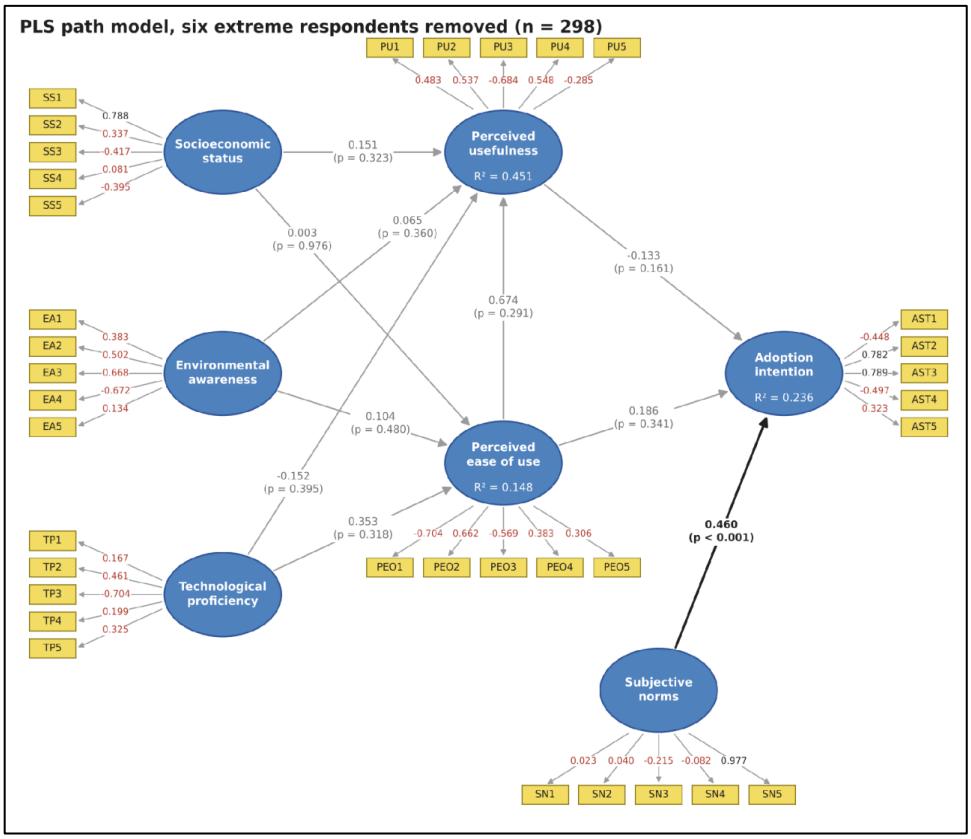


Figure 3: PLS Path Model, Six Extreme Respondents Removed (N = 298)

Table 4: Structural Estimates with and without the Six Extreme Respondents

Path	**β (N = 304)**	**95% CI**	***p***	**β (n = 298)**	**95% CI**	***p***
SS → PU	0.099	[0.048, 0.153]	<0.001	0.151	[−0.223, 0.327]	0.323
EA → PU	0.383	[0.231, 0.535]	<0.001	0.065	[−0.084, 0.188]	0.36
TP → PU	−0.112	[−0.209, −0.019]	0.026	−0.152	[−0.294, 0.328]	0.395
SS → PEOU	0.132	[0.063, 0.186]	<0.001	0.003	[−0.197, 0.163]	0.976
EA → PEOU	0.66	[0.542, 0.777]	<0.001	0.104	[−0.215, 0.265]	0.48
TP → PEOU	0.192	[0.069, 0.299]	0.001	0.353	[−0.436, 0.448]	0.318
PEOU → PU	0.613	[0.487, 0.733]	<0.001	0.674	[−0.749, 0.772]	0.291
PU → AI	0.224	[0.079, 0.357]	0.001	−0.133	[−0.324, 0.028]	0.161
PEOU → AI	0.146	[0.021, 0.287]	0.033	0.186	[−0.277, 0.339]	0.341
SN → AI	0.577	[0.432, 0.687]	<0.001	0.46	[0.333,	<0.001

					0.560]	
EA → PU → AI	0.086	[0.024, 0.149]	0.008	−0.009	[−0.038, 0.015]	0.537
EA → PEOU → AI	0.096	[0.014, 0.197]	0.039	0.019	[−0.040, 0.065]	0.454
R^2 (AI)	0.85			0.236		
Permutation-null R^2	0.042			0.068		

Note: 1,000 bootstrap resamples, percentile confidence intervals. The last two rows are indirect effects.

Bootstrap Blindness (P3)

None of the 1,000 bootstrap resamples excluded all six cases, consistent with the analytical probability of 0.2%. On average a resample contained 6.0 draws of the six cases, and the number of draws correlated 0.84 with the resamples R^2 (Table 5). Resamples with one draw had a mean R^2 of 0.49; those with eleven or more had a mean above 0.91. The bootstrap therefore estimates sampling variability around a result that the cluster creates, rather than testing whether the result exists without it. P3 is supported.

Table 5: R^2 of Adoption Intention by Number of Draws of the Six Cases in Each Bootstrap Resample

Draws of the six cases	**Resamples**	**Mean R^2 (AI)**	**Mean SN → AI**
1	20	0.487	0.434
2	40	0.663	0.503
3	102	0.744	0.529
4	129	0.795	0.554
5	159	0.83	0.568
6	162	0.852	0.576
7	138	0.871	0.587
8	85	0.883	0.589
9	73	0.894	0.592
10	34	0.901	0.59
11 or more	58	≥ 0.913	

Dose-Response (P4)

Removing the six cases one by one produced a smooth dose-response curve (Figure 4). Each removed case lowered R^2 for intention, from 0.850 with all six to 0.742 with three removed and 0.236 with all removed. Reliability fell in step: α for intention dropped from 0.866 to 0.093, and α for subjective norms from 0.888 to −0.596. The effect of EA on PEOU, the central path of earlier antecedent models, fell from 0.660 to 0.104. The removal order made little difference (ranges across subsets were narrow), so the result does not depend on any single respondent. P4 is supported.

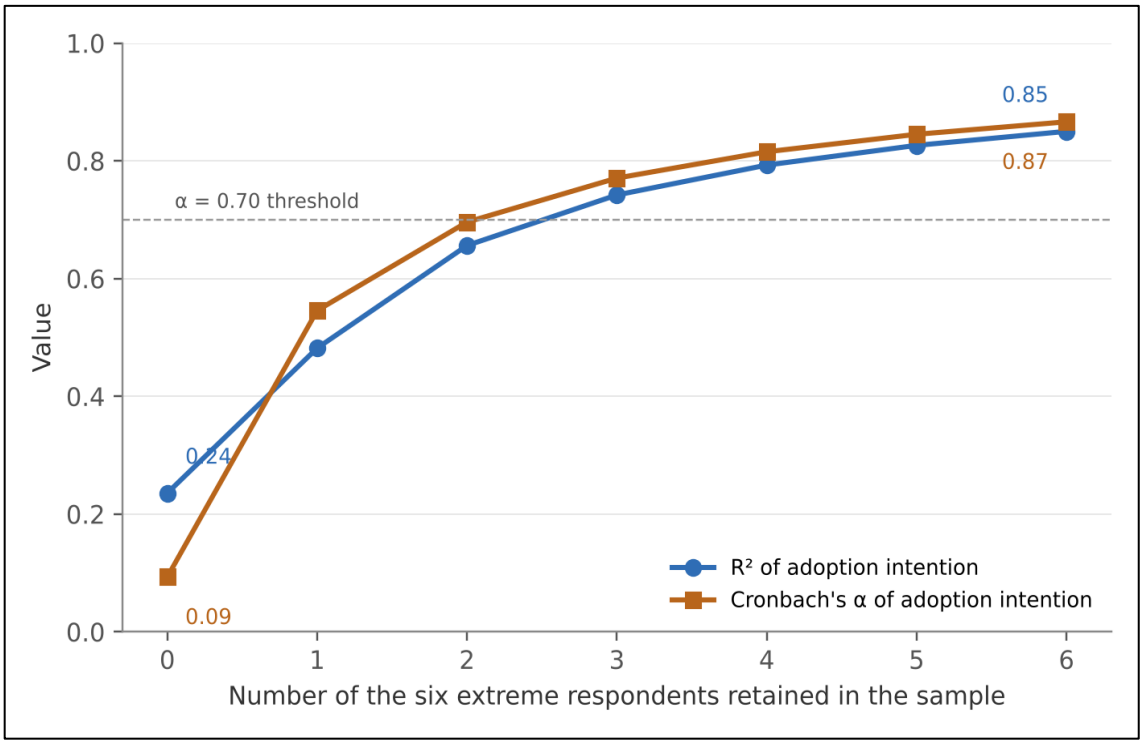


Figure 4: R^2 And Reliability of Adoption Intention as the Extreme Cluster is Removed

Advanced Extensions (P5)

Each extension produced publishable-looking results at N = 304 that disappeared without the six cases (Table 6). NCA identified every predictor as necessary for intention, with effect sizes of 0.46–0.72 ($p < 0.001$), because the only cases with low intention were the six. Without them, effect sizes fell to 0–0.11 (all $p \geq 0.17$). With standard calibration anchors, fsQCA assigned 98% of cases full membership in every set, leaving a single trivial solution. With anchors set within the agreement range, it produced a five-term solution (consistency 0.90, coverage 0.92), but random permutations of the outcome produced a solution of the same size and consistency in 100% of 200 runs. An education-based multigroup analysis found significant differences for PU → AI and TP → AI ($p = 0.026$ and 0.024), but all six extreme cases were in the lower-education group, MICOM compositional invariance failed, and the differences vanished at n = 298 (all $p \geq 0.40$).

IPMA placed every construct at 87–95 on the 0–100 performance scale at N = 304 and became undefined at n = 298, where negative weights pushed performance values above 100. PLSpredict gave Q^2predict values of 0.49–0.61 at N = 304, falling to −0.02–0.20 at n = 298. P5 is supported: the extensions did not provide independent corroboration, because they shared the same source of variance.

Table 6: Advanced PLS-SEM Extensions with and without the Six Extreme Respondents

Extension	**N = 304**	**n = 298**	**Diagnostic that signaled the problem**
NCA (CE-FDH)	d = 0.46–0.72, all p < 0.001	d = 0.00–0.11, all p ≥ 0.17	Only six cases below the ceiling
fsQCA (within-agreement calibration)	5-term solution, cons. 0.90, cov. 0.92	Same solution	Permuted outcome reproduced it in 100% of runs
MICOM + MGA (education)	PU → AI and TP → AI differ (p = 0.026, 0.024)	No differences (p ≥ 0.40)	Compositional invariance failed
IPMA	Performance 87.6–94.5 for all constructs	Undefined (performance >	Negative indicator weights

		100)	
PLSpredict	Q^2predict 0.49–0.61; PLS beats LM on 2/5 items	Q^2predict −0.02 to 0.20	Weak benchmark comparison

Note: NCA, fsQCA and IPMA were computed on construct scores of the focal constructs; MGA used a direct-effects specification (PU, PEOU, SN, SS and TP → AI).

Detection (P6)

The univariate careless-response indices did not flag the six cases. Their longest runs of identical answers (4–8 items) were shorter than the median of the other respondents (11), and their intra-individual response variability (0.44–0.50) was almost identical to the others' median (0.43), because they varied between 1 and 2 just as the majority varied between 4 and 5. By contrast, the Mahalanobis distance on the seven construct means flagged all six at $p < 0.001$ (D^2 = 48.9–53.2, against a median of 5.2), along with three other respondents. P6 is supported: in ceiling-affected data, a multivariate outlier screen followed by case-removal sensitivity analysis is the effective safeguard.

Monte Carlo Simulation

The simulation shows that the pattern is not peculiar to this dataset (Figure 5; Table 7). In pure noise, with every item shuffled, the model correctly showed nothing: mean R^2 was 0.07 and mean α was −0.01. Adding just three synthetic all-disagree respondents (1% of the sample) raised R^2 to 0.71, α to 0.73 and the share of HTMT ratios above 0.90 to 99.5%. With six cases, R^2 reached 0.83, α 0.84 and all three paths into intention were significant in 99.5% of replications, all from data containing no relationships at all. The real data behaved almost identically once three or more extreme cases were added, which means that conventional statistics cannot distinguish a real model from noise in this situation.

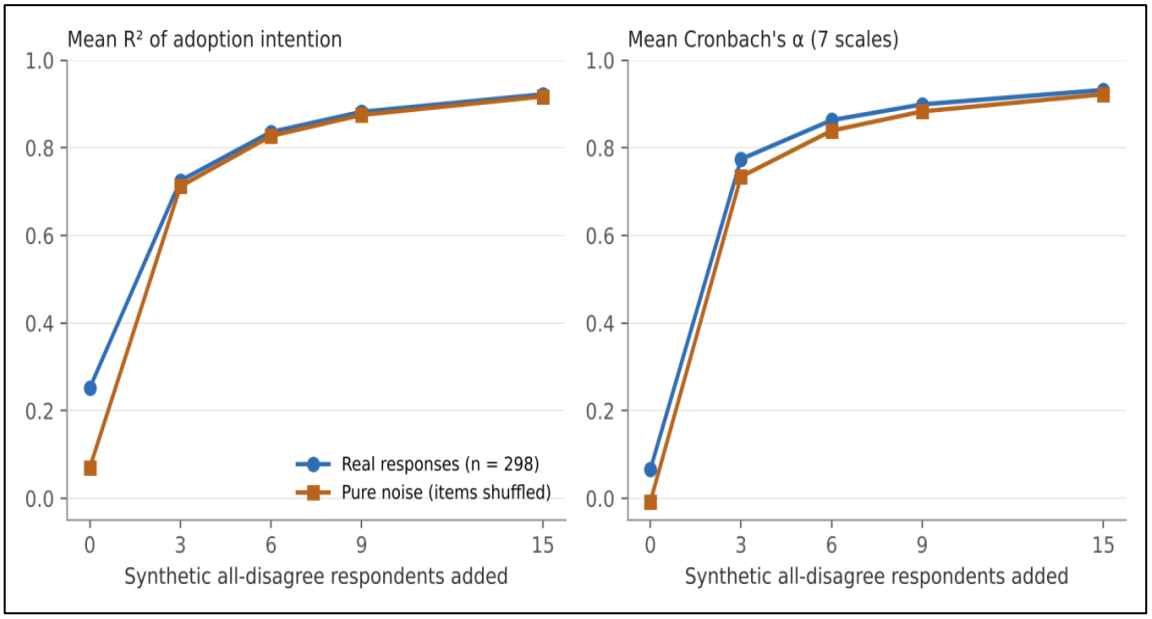


Figure 5. Simulation: Synthetic All-Disagree Respondents Added to Real and Pure-Noise Data

Table 7: Simulation Results (200 Replications Per Cell)

Data	Extreme cases added	Mean R^2 (AI)	Mean α	Share HTMT > 0.90	All AI paths significant
Pure noise	0	0.07	−0.008	4.20%	15.00%
Pure noise	3	0.712	0.734	99.50%	98.00%
Pure noise	6	0.827	0.839	100.00%	99.50%
Pure noise	9	0.875	0.883	100.00%	100.00%

Pure noise	15	0.917	0.922	100.00%	100.00%
Real responses	0	0.252	0.066	7.10%	
Real responses	3	0.725	0.774	65.30%	
Real responses	6	0.836	0.863	70.00%	
Real responses	9	0.882	0.899	71.50%	
Real responses	15	0.922	0.932	80.00%	

DISCUSSION

The central finding is that six respondents, 2% of the sample, generated almost every result that a conventional PLS-SEM report would contain: high reliability, convergent validity, significant direct and indirect effects, high R^2 and Q^2predict, necessary conditions, configurational solutions, group differences and managerial priority maps. Because all these outputs are built from the same correlation matrix, they agreed with one another, and their agreement looked like triangulation. In fact, it was one piece of evidence counted many times.

The mechanism is straightforward. When 98% of respondents score near the top of the scale, the covariance within that group is close to zero. A small group at the opposite end of the distribution then creates a single axis of variance along which all constructs are aligned. That axis inflates reliability, erodes discriminant validity (high HTMT), produces the dominant first factor in Harman's test and a high R^2, and makes the extreme cases the only observations below the NCA ceiling and outside the fsQCA membership sets. The pattern connects four established concerns: response-style bias (Baumgartner & Steenkamp, 2001), careless responding (Ward & Meade, 2023), influential cases (Aguinis et al., 2013) and the tendency of PLS-SEM to capitalise on chance (Rönkkö et al., 2016).

The diagnostics routinely used in this field are poorly suited to this situation. Bootstrap confidence intervals treat the sample as a stand-in for the population, so resamples almost always contain the influential cases. HTMT and full-collinearity VIFs did signal a problem, but they attributed it to construct overlap or method bias rather than to a small group of respondents. The failed MICOM test in the education comparison was the only routine diagnostic that pointed to the true source. Standard careless-responding indices flagged none of the six cases, because their answers varied within the disagreement range just as the majority's varied within the agreement range.

When it comes to tourism technology adoption, research usually focuses on desirable innovations that are environmentally sustainable, employing positive scales. Moreover, in collectivist settings, where agreement can be a normative response style (Baumgartner & Steenkamp, 2001), ceiling effects should be expected rather than treated as exceptional. Papers that have reported an R^2 of 0.80 or higher, α of 0.90 or higher, and HTMT of 1 or higher in such environments should be treated with caution unless sensitivity to extreme respondents has been proven. The same caution should be exercised in the growing use of necessary condition analysis (NCA) and fuzzy-set QCA (fsQCA) alongside PLS-SEM since both approaches pay attention to the extremes, which is where a small number of extreme respondents is located. The remedy is not to abandon PLS-SEM but to demonstrate that substantive conclusions survive the removal of a small number of extreme respondents.

Whether such respondents should be retained depends on who they are. If they are careless responders, their answers carry no information about the constructs and the full-sample results are artefacts. If they are genuine rejecters, they form a real but very small segment whose attitudes differ in level, not in structure, from the majority; the model then describes the gap between two groups rather than relationships within the population. In neither case do the full-sample estimates describe how the constructs relate among the 98% of respondents who agree. The recommendation of this study is therefore not to delete such cases, but to report results with and without them, to state which interpretation is adopted and why, and to collect the information needed to tell the two apart, such as completion timestamps, attention checks and reverse-keyed items.

CONCLUSION

Using real tourism data and simulation, this study shows that a cluster of six uniformly negative respondents can produce a PLS-SEM model that passes every common test yet carries virtually no information about the remaining respondents. Without the cluster, only a modest effect of subjective norms on adoption intention remains ($\beta = 0.460$; $R^2 = 0.236$). Because the cluster inflates all statistics simultaneously, multi-method corroboration offers little protection. The study has limitations.

1. One case dataset. The empirical results come from a single survey. The simulation suggests the pattern generalizes to other ceiling-affected data, but replication across tourism datasets with different scale formats and response distributions is needed.

2. Unknown status of the six cases. Without timestamps or attention checks it cannot be determined whether the six were careless or genuine rejecters. The study does not claim they should be deleted; it shows that conclusions must be reported both ways.

3. Estimation software. Estimates were produced with a Python implementation of the PLS algorithm that reproduces SmartPLS defaults. Small differences from commercial software are possible; the R script in Appendix A allows independent replication.

4. Scope of remedies. The study focuses on detection and reporting. Future research should evaluate design remedies, such as balanced or reverse-keyed items, wider response scales, attention checks and forced-choice formats, and estimation remedies such as robust or weighted PLS.

ACKNOWLEDGEMENT

This publication was supported by Universiti Teknikal Malaysia Melaka (UTeM) under the Journal Publication Fee Initiative 2026. The authors would also acknowledge support from the Faculty of Technology Management and Technopreneurship.

ETHICS STATEMENT

This study received ethical approval from the UTeM Research Ethics Committee (REAC), Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, under Reference Number REAA/2026/06-01. The study was conducted in accordance with the ethical principles and research governance requirements established by Universiti Teknikal Malaysia Melaka and applicable national research ethics guidelines.

REFERENCES

1. Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848
2. Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management*, 30(1), 514–538. https://doi.org/10.1108/IJCHM-10-2016-0568
3. Azmi, F. R., Abdullah, A., Musa, H., & Wan Mahmood, W. H. (2020). Perception of food manufacturers towards adoption of halal food supply chain in Malaysia. *Journal of Islamic Marketing*, 11(3), 571–589. https://doi.org/10.1108/JIMA-12-2018-0236
4. Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156. https://doi.org/10.1509/jmkr.38.2.143.18840
5. Becker, J.-M., Cheah, J.-H., Gholamzade, R., Ringle, C. M., & Sarstedt, M. (2023). PLS-SEM's most wanted guidance. *International Journal of Contemporary Hospitality Management*, 35(1), 321–346. https://doi.org/10.1108/IJCHM-04-2022-0474

6. Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19.
7. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
8. do Valle, P. O., & Assaker, G. (2016). Using partial least squares structural equation modeling in tourism research: A review of past research and recommendations for future applications. *Journal of Travel Research*, 55(6), 695–708. https://doi.org/10.1177/0047287515569779
9. Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods*, 19(1), 10–52. https://doi.org/10.1177/1094428115584005
10. Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall.
11. Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24.
12. 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.
13. Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review*, 33(3), 405–431.
14. Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10.
15. Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455.
16. Musa, H., Abdullah, A. R., Othman, M. N., Rashid, N., & Azmi, F. R. (2025). Rethinking the technology acceptance model: A structural analysis of sustainable technology adoption in Malaysian tourism. *International Journal of Research and Innovation in Social Science*, 9(9), 6125–6139. https://doi.org/10.47772/IJRISS.2025.909000500
17. Musa, H., Li, S. C. H., Abas, Z. A., & Mohamad, N. (2016). Adoption factor of mobile marketing: The case of small medium enterprises in Malaysia. *International Review of Management and Marketing*, 6(7S), 112–115.
18. Mustafa, S. K., Musa, H., Abdullah, A. R., & Azmi, F. R. (2017). The influences of consumer preference characteristics towards green product. *International Journal of Human and Technology Interaction*, 1(1), 43–50.
19. Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, 58, 102310. https://doi.org/10.1016/j.ijinfomgt.2021.102310
20. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
21. Rasoolimanesh, S. M., Ringle, C. M., Sarstedt, M., & Olya, H. (2021). The combined use of symmetric and asymmetric approaches: Partial least squares-structural equation modeling and fuzzy-set qualitative comparative analysis. *International Journal of Contemporary Hospitality Management*, 33(5), 1571–1592. https://doi.org/10.1108/IJCHM-10-2020-1164
22. Richter, N. F., Schubring, S., Hauff, S., Ringle, C. M., & Sarstedt, M. (2020). When predictors of outcomes are necessary: Guidelines for the combined use of PLS-SEM and NCA. *Industrial Management & Data Systems*, 120(12), 2243–2267. https://doi.org/10.1108/IMDS-11-2019-0638
23. Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS-SEM results: The importance-performance map analysis. *Industrial Management & Data Systems*, 116(9), 1865–1886. https://doi.org/10.1108/IMDS-10-2015-0449
24. Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, 47–48, 9–27. https://doi.org/10.1016/j.jom.2016.05.002
25. Safri, N., & Musa, H. (2022). The adoption factors of low loss microwave transmission glass among green buildings in Malaysia. *Journal of Positive School Psychology*, 6(3), 4878–4897.

26. Sarstedt, M., Ringle, C. M., Cheah, J.-H., Ting, H., Moisescu, O. I., & Radomir, L. (2020). Structural model robustness checks in PLS-SEM. *Tourism Economics*, 26(4), 531–554.

https://doi.org/10.1177/1354816618823921
27. Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347.
28. Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology*, 74, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007