

Structural equation modelling in R: Comparing *lavaan* and *semnr* packages

Methodological Innovations
2026, Vol. 19(2) 118–132
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20597991261445104
journals.sagepub.com/home/mio



Isaac Osei¹ , Chettupally Anil Carie¹, Satish Anamalamudi¹,
Dennis Opoku Boadu² and Michael Kwakye³

Abstract

This paper compares two popular R packages for structural equation modelling (SEM) – *lavaan* and *semnr* – to help researchers understand not only how they work, but also when each is most appropriate. Although both tools allow users to estimate the same structural models, they are grounded in different methodological traditions and are designed to support different research goals. Using an identical model, we estimated results with both covariance-based SEM (*lavaan*) and variance-based SEM (*semnr*) and compared their outputs, including model specification syntax, evaluation criteria, and reporting conventions. The results show that both approaches lead to substantively similar conclusions regarding the relationships between constructs, while differing in emphasis: *lavaan* provides richer global model-fit diagnostics, whereas *semnr* places greater emphasis on prediction-oriented assessment and convenient access to latent variable scores. The contribution of this study lies in its practical, hands-on demonstration rather than in a theoretical or simulation-based comparison. The findings reinforce that there is no universally “better” SEM approach; instead, methodological choice should be guided by the research objective. Researchers focussed on theory testing may benefit more from *lavaan*, while those prioritising prediction or exploratory analysis may find *semnr* more suitable. Ultimately, considering both perspectives can support more transparent, robust, and methodologically appropriate SEM applications.

Keywords

SEM, *lavaan*, *semnr*, PLS-SEM, CB-SEM

Introduction

Structural equation modelling (SEM) has become one of the most powerful tools in the researcher’s toolkit, particularly in disciplines such as behavioural sciences, psychology, education, information systems, and marketing (Guenther et al., 2023; Hair et al., 2021, 2022; Sarstedt et al., 2022). Its appeal lies in its ability to examine complex theoretical models by combining measurement models – linking latent variables to their observed indicators – with structural models that map out the relationships between those latent variables (Hair et al., 2021; Kline, 2016). In other words, SEM allows researchers to go beyond simple correlations and regressions, enabling them to test entire theories in a single, cohesive framework (Almeida, 2024; Chin, 1998; Kline, 2023).

Within the *R*, an open-source software (R Core Team, 2025) ecosystem, researchers are fortunate to have access to several robust packages for SEM. Among them, *lavaan* (Rosseel, 2012; Rosseel et al., 2025) and *semnr* (Ray et al.,

¹Department of Computer Science and Engineering, SRM University-AP, Amaravati, Andhra Pradesh, India

²Department of Computer Science, University of Ghana, Legon, Accra, Ghana

³Department of Information Technology and Systems, Ghana Christian University College, Accra, Ghana

Corresponding author:

Dennis Opoku Boadu, Department of Computer Science, University of Ghana, Legon, Accra, 00233, Ghana.

Email: doboadu@st.ug.edu.gh



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

2025) stand out as two of the most widely adopted. Although they serve the same overarching purpose, the philosophy behind their design and the problems they address differ significantly.

The *lavaan* package is the workhorse of covariance-based SEM (CB-SEM; Kline, 2016). It provides an extensive suite of functions for model specification, parameter estimation, model fit evaluation, and hypothesis testing. It is ideal for researchers focussed on theory confirmation and model fit assessment under the assumption of multivariate normality. On the other hand, *semnr* was developed to make partial least squares SEM (PLS-SEM; Hair et al., 2021), a variance-based approach, more accessible to R users. PLS-SEM is particularly useful when the goal is prediction rather than theory testing, or when researchers must deal with small sample sizes, formative measurement models, or non-normal data.

While numerous studies have discussed the conceptual differences between CB-SEM and PLS-SEM (eg, Dash and Paul, 2021; Rigdon et al., 2017; Sarstedt et al., 2016; Schuberth et al., 2023), there is limited practical guidance on how the two approaches compare when applied to the same dataset using R. Researchers often face uncertainty about which package to choose, especially when theoretical models could be analysed using either method (Sakaria et al., 2023; Vuković, 2024). Therefore, a systematic, hands-on comparison is both timely and valuable for applied researchers seeking clarity in methodological selection.

The aim of this paper is to offer a practical, side-by-side comparison of *lavaan* and *semnr*. We demonstrate how the same model can be specified and estimated using both packages, highlighting differences in syntax, output, and interpretation. By doing so, we hope to provide researchers with clear guidance on selecting the approach that best matches their data characteristics, research objectives, and theoretical goals.

The study is guided by the following research questions:

1. **RQ1:** Using the same dataset, do *semnr* and *lavaan* produce similar results?
2. **RQ2:** Which package offers greater flexibility and ease of use?
3. **RQ3:** Can the two packages be used complementarily?

By addressing these questions, this study aims to contribute to the methodological literature on when and how to appropriately choose between PLS-SEM and CB-SEM approaches.

The remainder of this paper is organised as follows. Section 2 reviews the theoretical foundations of CB-SEM and PLS-SEM. Section 3 describes the dataset and the comparative model, outlining the implementation steps for both *lavaan* and *semnr* while highlighting key syntactical and estimation differences. Section 4 presents and compares the empirical results, and Section 5 interprets the findings, discussing their

methodological implications, practical recommendations, and directions for future research. Finally, Section 6 concludes the paper by summarising the key insights.

Literature review

Structural Equation Modelling (SEM) has evolved over several decades into one of the most versatile tools for empirical research. Its origins can be traced to the combination of path analysis (Hair et al., 2017; McDonald, 1996; Wright, 1921) and factor analysis (Spearman, 1904), resulting in a methodology that allows researchers to assess both measurement validity and structural relationships in a single statistical framework (Bollen, 1989; Bollen and Diamantopoulos, 2017; Kline, 2023; Lohmöller, 1989; Vinzi et al., 2010; Wold, 1975). Over the years, SEM has been widely adopted in disciplines such as psychology, marketing, and information systems, where researchers often deal with latent variables or constructs that cannot be measured directly.

At its core, SEM distinguishes between latent variables, which represent theoretical constructs that are not directly observable, and manifest (observed) variables, which serve as empirical indicators of those constructs. These relationships are formalised through a measurement model, which links latent variables to their observed indicators, and a structural model, which specifies the hypothesised directional relationships (paths) among latent variables (Hair et al., 2021, 2022; Kline, 2023). By integrating these two components within a single analytical framework, SEM enables researchers to simultaneously assess measurement validity and test theoretical relationships.

Covariance-based SEM (CB-SEM; Hair et al., 2017, 2025; Reinartz et al., 2009) is traditionally used for theory testing, focussing on how well a proposed model fits the observed data (Bollen, 1989; Kline, 2016). It has been the dominant approach to SEM. CB-SEM aims to reproduce the observed covariance matrix as closely as possible, placing strong emphasis on model fit and theory confirmation (Kline, 2016, 2023). Its popularity is partly due to software such as LISREL, AMOS, and Mplus, and more recently, the open-source *lavaan* package (Rosseel, 2012; Rosseel et al., 2025) in R. *Lavaan* has become the *de facto* standard for researchers seeking a free, powerful, and flexible implementation of CB-SEM. It supports a wide range of models – confirmatory factor analysis (CFA), mediation, moderation, multigroup analysis, and more – while offering robust options for model fit evaluation through indices such as CFI, TLI, RMSEA, and SRMR.

In contrast, variance-based SEM (PLS-SEM; Cepeda et al., 2024; Hair et al., 2022; Sarstedt et al., 2017; Schuberth et al., 2025) has grown in popularity for its predictive focus and flexibility in handling complex models, formative constructs, small sample sizes, and non-normal data (Hair et al., 2022). PLS-SEM does not attempt to reproduce the covariance matrix but rather maximises the explained variance of

endogenous constructs. This makes it especially suitable for exploratory studies or prediction-oriented research (Chin, 1998; Goktas and Dirsehan, 2025; Richter and Tudoran, 2024). Traditionally, PLS-SEM has been associated with proprietary software such as SmartPLS and WarpPLS, but with the advent of *seminr*, researchers can now perform PLS-SEM directly within R, enabling reproducibility and seamless integration with other statistical workflows. *Seminr* (Ray et al., 2021, 2025) implements modern PLS-SEM algorithms and is popular for its intuitive model specification using *constructs()* and *relationships()* functions.

Recent methodological works (Gudergan et al., 2025; Hair et al., 2021, 2022; Sakaria et al., 2023) have emphasised that CB-SEM and PLS-SEM are complementary rather than competing approaches. CB-SEM is recommended when the research objective is theory confirmation and model fit evaluation, while PLS-SEM is recommended when the focus is on prediction, theory development, or when data do not meet CB-SEM assumptions (Hair and Alamer, 2022; Hair et al., 2019; Sharma et al., 2024).

Despite the availability of both *lavaan* and *seminr*, there is still limited academic discussion comparing the two tools in a systematic way. Most studies focus on either CB-SEM or PLS-SEM exclusively, rarely demonstrating how a researcher might specify and analyse the same model using both approaches. A few comparative studies (eg, Dash and Paul, 2021; Reinartz et al., 2009; Sarstedt et al., 2024) have examined the conceptual differences between CB-SEM and PLS-SEM, but practical, hands-on guides remain scarce, especially for R users who must decide which package best suits their data characteristics and research objectives.

This gap motivates the present study. By providing a side-by-side comparison of *lavaan* and *seminr*, this paper contributes to the methodological literature by clarifying when each package is most appropriate, illustrating their syntax and outputs, and offering guidance for applied researchers.

Methodology

The purpose of this study is to provide a hands-on, side-by-side comparison of the *lavaan* and *seminr* packages in R. To achieve this, we adopt a demonstration-based research design, specifying and estimating the same structural equation model (SEM) using both packages and then comparing the syntax, estimation process, and results. This section outlines the model, data, procedures, and evaluation criteria.

Model specification

A hypothetical SEM was constructed to include both a measurement model and a structural model, reflecting typical use cases in the social sciences. The model consists of six latent variables (constructs) measured by multiple reflective indicators, along with structural paths linking the constructs to test hypothesised relationships. This design ensures that the

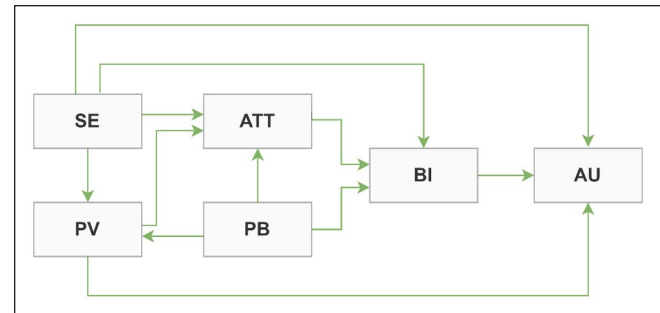


Figure 1. Conceptual framework of the proposed model.

Source: Reproduced from authors' previous work.

example is sufficiently complex to demonstrate both covariance-based SEM (CB-SEM) and variance-based SEM (PLS-SEM) capabilities.

- **Measurement model:** Each construct is measured using three observed variables (indicators). The population measurement model was designed to reflect commonly recommended reliability benchmarks in SEM research. Standardised loadings were specified above 0.70, internal consistency was targeted above 0.70 (Cronbach's alpha and composite reliability), and Average Variance Extracted (AVE) was specified above 0.50. All constructs were modelled reflectively with independent measurement errors. These design choices were implemented to create a pedagogically sound dataset for comparison purposes.
- **Structural model:** The model specifies direct relationships between the exogenous construct and at least one endogenous construct, allowing us to estimate path coefficients and assess explained variance (R^2 values).

The proposed conceptual model (Figure 1) posits a series of hypothesised relationships among the study constructs. Specifically, social effects (SE) and perceived benefits (PB) are expected to jointly predict perceived value (PV); SE, PB, and PV are hypothesised to predict attitude (ATT); ATT, PB, and SE are proposed to influence behavioural intention (BI); and finally, BI, SE, and PV are anticipated to predict actual use (AU). This model was specified in both *lavaan* and *seminr*.

```

lavaan syntax
model <- '
  # Measurement model
  SE =~ se1 + se2 + se3
  PB =~ pb1 + pb2 + pb3
  PV =~ pv1 + pv2 + pv3
  ATT =~ att1 + att2 + att3
  BI =~ bi1 + bi2 + bi3
  AU =~ au1 + au2 + au3
  # Structural model

```

```

PV ~ SE + PB
ATT ~ SE + PV + PB
BI ~ ATT + PB + SE
AU ~ BI + SE + PV
,
lavaan_model <- sem(model, data = df,
meanstructure = TRUE)

seminr syntax
library(seminr)
# Measurement model
measurement <- constructs(
  composite("SE", multi_items("se", 1:3)),
  composite("PB", multi_items("pb", 1:3)),
  composite("PV", multi_items("pv", 1:3)),
  composite("ATT", multi_items("att", 1:3)),
  composite("BI", multi_items("bi", 1:3)),
  composite("AU", multi_items("au", 1:3))
)
# Structural model
structural <- relationships(
  paths(from = c("SE", "PB"), to = "PV"),
  paths(from = c("SE", "PB", "PV"), to =
"ATT"),
  paths(from = c("ATT", "PB", "SE"), to =
"BI"),
  paths(from = c("BI", "SE", "PV"), to =
"AU")
)
seminr_model <- estimate_pls(data = df,
measurement_model = measurement, struc-
tural_model = structural)

```

Data

For the purposes of demonstration, a synthetic dataset was generated with realistic parameter values and sample size ($N=500$), ensuring that model identification criteria are satisfied. Synthetic data allows for full control of measurement error, construct correlations, and path strengths, and avoids concerns of data privacy or domain-specific bias. To ensure full reproducibility, the data were generated using a fixed random seed in R (version 4.5.1). The dataset was simulated using the *simulateData()* function in *lavaan*, ensuring consistency between the population model and the subsequent analyses. All observed variables (indicators) were generated as continuous measures from the underlying population model and subsequently transformed into Likert-type scales for analysis. For this comparison, we use a dataset measuring six latent variables: Social effects (SE), perceived benefits (PB), perceived value (PV), attitude (ATT), behavioural intention (BI), and actual use (AU). Each construct is measured using three reflective indicators. This discretisation

step was applied for illustrative purposes and was not intended to evaluate estimator performance under ordinal data conditions. Because the dataset is synthetic, results are presented solely for illustrative comparison of software outputs rather than for substantive inference. The complete R code used to generate the synthetic dataset, including the random seed, is provided in Appendix A to facilitate full reproducibility.

Analysis in lavaan (CB-SEM)

The model was first estimated using the *lavaan* package (Rosseel, 2012; Rosseel et al., 2025), which applies maximum likelihood (ML) estimation under the assumption of multivariate normality. All analyses were conducted using *lavaan* (version 0.6-20) to ensure reproducibility. The following steps were performed:

1. **Model specification:** The measurement and structural model were defined using *lavaan*'s formula-like syntax.
2. **Parameter estimation:** ML estimation was applied to obtain factor loadings, path coefficients, and covariances.
3. **Model fit assessment:** Global model fit was assessed using CFI, TLI, NFI, RMSEA, and SRMR indices.
4. **Significance testing:** Standard errors and p -values were reported to evaluate statistical significance of the parameters.

Analysis in seminr (PLS-SEM)

Next, the model was specified and estimated using *seminr* (Ray et al., 2021, 2025), which implements the Partial Least Squares (PLS) algorithm. The analyses were performed using *seminr* (version 2.3.7), consistent with current methodological recommendations. The steps included:

1. **Model specification:** Constructs, indicators, and structural paths were defined using *seminr*'s declarative syntax (*constructs()*, *composite()*, *paths()*, *relationships()*).
2. **Model estimation:** PLS estimation was performed to maximise the explained variance of endogenous constructs.
3. **Quality assessment:** Reliability (Cronbach's alpha and composite reliability), convergent validity (AVE), and discriminant validity (Fornell-Larcker and HTMT criteria) were assessed.
4. **Bootstrapping:** A bootstrapping procedure with 5000 resamples was used to compute standard errors, t -statistics, and significance levels for path coefficients.

Table 1. Measurement model results (*lavaan*).

Construct	Indicator	Loading	Std err	Z	p-value	CI lower	CI upper
SE	se1	0.797	0.023	35.214	< 0.001	0.753	0.841
	se2	0.864	0.020	43.087	< 0.001	0.825	0.904
	se3	0.726	0.026	27.940	< 0.001	0.675	0.777
PB	pb1	0.729	0.026	28.072	< 0.001	0.678	0.780
	pb2	0.830	0.021	39.300	< 0.001	0.789	0.872
	pb3	0.805	0.022	36.238	< 0.001	0.762	0.849
PV	pv1	0.738	0.023	31.495	< 0.001	0.692	0.784
	pv2	0.887	0.015	59.462	< 0.001	0.858	0.917
	pv3	0.851	0.017	50.809	< 0.001	0.818	0.884
ATT	att1	0.826	0.020	42.242	< 0.001	0.788	0.865
	att2	0.806	0.021	38.976	< 0.001	0.765	0.846
	att3	0.747	0.024	31.072	< 0.001	0.699	0.794
BI	bi1	0.836	0.017	47.943	< 0.001	0.801	0.870
	bi2	0.870	0.016	56.043	< 0.001	0.840	0.900
	bi3	0.822	0.018	45.091	< 0.001	0.786	0.858
AU	au1	0.865	0.015	58.804	< 0.001	0.836	0.893
	au2	0.915	0.012	75.433	< 0.001	0.891	0.939
	au3	0.845	0.016	53.338	< 0.001	0.814	0.876

Source: Authors' own compilation.

Note. Std err: standard error; Z: test statistics; CI: confidence interval.

Comparison criteria

The results from *lavaan* and *semnr* were compared on the following dimensions:

- **Syntax and ease of use:** The intuitiveness, readability, and conciseness of model specification.
- **Output:** The type and breadth of parameter estimates, model fit indices, and reliability and validity measures produced by each package.
- **Interpretation:** The extent to which the outputs support theory confirmation in CB-SEM (*lavaan*) versus prediction-oriented assessment in PLS-SEM (*semnr*).
- **Performance:** A descriptive comparison of estimated path coefficients, factor loadings, and coefficients of determination (R^2) for endogenous constructs.

The goal of this comparison is not to determine the superiority of one approach over the other, but rather to illustrate their respective strengths, trade-offs, and practical implications when applied to the same model and dataset. Accordingly, differences observed in the results are interpreted in light of the underlying estimation philosophies and output conventions of CB-SEM and PLS-SEM, rather than as evidence of methodological dominance.

Results

This section presents the outcomes of estimating the same structural equation model (SEM) using *lavaan* (CB-SEM)

and *semnr* (PLS-SEM). The results are organised into three parts: measurement model assessment, structural model results, and a comparative summary.

Measurement model results

Lavaan (CB-SEM). As shown in Table 1, the confirmatory factor analysis (CFA) results indicate that all factor loadings were statistically significant ($p < 0.001$) and exceeded the recommended threshold of 0.70–0.708 (Cepeda et al., 2024; Hair et al., 2014, 2019), thereby establishing strong indicator reliability. Composite reliability (CR) and average variance extracted (AVE) were not automatically generated because *lavaan* does not compute these metrics by default. However, they can be derived manually or obtained using supplementary packages such as *semTools* (Jorgensen et al., 2025). These results suggest that the measurement model fits the data well and adequately represents the underlying constructs.

Seminr (PLS-SEM). The PLS-SEM results (Table 2) provided strong support for the reliability and validity of the measurement model. All indicator loadings exceeded the recommended 0.708 threshold, confirming indicator reliability. Multicollinearity was not a concern, as all VIF values ranged from 1.69 to 3.54, well below the conservative cutoff of 5. Internal consistency was robust, with Cronbach's alpha values between 0.829 and 0.906 and composite reliability (rhoC) values ranging from 0.897 to 0.941. Convergent validity was also established, as all constructs achieved AVE values between 0.745 and 0.842, comfortably above the 0.50

Table 2. Measurement model results (seminr).

Construct	Indicator	Loading	VIF	alpha	rhoC	AVE	rhoA
SE	se1	0.867	2.052	0.836	0.902	0.754	0.837
	se2	0.894	2.285				
	se3	0.841	1.741				
PB	pb1	0.842	1.699	0.829	0.897	0.745	0.829
	pb2	0.878	2.179				
	pb3	0.868	1.985				
PV	pv1	0.839	1.839	0.862	0.916	0.785	0.870
	pv2	0.915	2.704				
	pv3	0.902	2.564				
ATT	att1	0.886	2.207	0.833	0.900	0.750	0.833
	att2	0.871	2.080				
	att3	0.839	1.694				
BI	bi1	0.893	2.478	0.879	0.925	0.805	0.880
	bi2	0.911	2.719				
	bi3	0.888	2.214				
AU	au1	0.916	2.857	0.906	0.941	0.842	0.909
	au2	0.934	3.541				
	au3	0.903	2.770				

Source: Authors' own compilation.

Note. VIF: variance inflation factor; alpha: Cronbach's alpha; rhoC: composite reliability; AVE: average variance extracted; rhoA: consistent reliability coefficient.

Table 3. Heterotrait-monotrait (HTMT) ratio.

Construct	SE	PB	PV	ATT	BI	AU
SE						
PB	0.500					
PV	0.508	0.596				
ATT	0.380	0.556	0.771			
BI	0.434	0.554	0.590	0.700		
AU	0.409	0.434	0.455	0.472	0.624	

Source: Authors' own compilation.

benchmark (Hair et al., 2021, 2022). The rhoA values (0.829–0.909) further supported construct reliability.

Discriminant validity was further examined using both the HTMT criterion (Henseler et al., 2015) and the Fornell-Larcker criterion (Fornell and Larcker, 1981). As shown in Table 3, all HTMT ratios ranged between 0.380 and 0.771 – well below the conservative threshold of 0.85 (Henseler et al., 2015) – indicating strong discriminant validity. The Fornell-Larcker results (Table 4) also supported this conclusion, with the square roots of AVE (diagonal values) exceeding the correlations among constructs in every case. Together, these results confirm that each construct in the model is empirically distinct and measures a unique conceptual domain.

Structural model results

Lavaan (CB-SEM). Table 5 summarises the structural model estimates. Most hypothesised paths were significant and in the expected directions. Perceived value (PV) was strongly

Table 4. Fornell-larcker criterion.

Construct	SE	PB	PV	ATT	BI	AU
SE	0.868					
PB	0.416	0.863				
PV	0.432	0.507	0.886			
ATT	0.318	0.462	0.653	0.866		
BI	0.372	0.473	0.516	0.599	0.897	
AU	0.356	0.378	0.409	0.412	0.559	0.918

Source: Authors' own compilation.

Note. Bolded values = $\sqrt{\text{AVE}}$; Unbolded values = inter-construct correlations.

predicted by social effects (SE; $\beta=0.276$, $p<0.001$) and perceived benefits (PB; $\beta=0.443$, $p<0.001$), explaining 39.3% of its variance. Attitude (ATT) was mainly driven by PV ($\beta=0.669$, $p<0.001$), with a total R^2 of 57.8%; the path from SE to ATT was the only non-significant relationship. Behavioural intention (BI) was significantly influenced by ATT, PB, and SE ($R^2=52.5\%$). Actual use (AU) was explained by BI, SE, and PV ($R^2=41.0\%$). Overall, the model demonstrates strong explanatory power across constructs.

The global fit indices demonstrated excellent overall model fit (Table 6), which also reports the standard benchmark thresholds widely recommended in SEM literature (Hair et al., 2019; Hu and Bentler, 1998; Kline, 2016). Although the Chi-square test was statistically significant, this result is expected given the sample size ($N=500$) and is therefore interpreted alongside other fit indices, consistent

Table 5. Hypotheses testing using *Lavaan*.

Path	β	R^2	SE	Z	p-value	CI lower	CI upper	Decision
SE→PV	0.276	0.393	0.051	5.438	<0.001	0.177	0.376	Supported
PB→PV	0.443	0.393	0.049	9.048	<0.001	0.347	0.539	Supported
SE→ATT	-0.044	0.578	0.050	-0.881	0.378	-0.143	0.054	Rejected
PV→ATT	0.669	0.578	0.047	14.123	<0.001	0.576	0.762	Supported
PB→ATT	0.174	0.578	0.055	3.153	0.002	0.066	0.282	Supported
ATT→BI	0.531	0.525	0.046	11.523	<0.001	0.441	0.622	Supported
PB→BI	0.203	0.525	0.055	3.688	<0.001	0.095	0.310	Supported
SE→BI	0.119	0.525	0.048	2.480	0.013	0.025	0.213	Supported
BI→AU	0.489	0.410	0.049	9.889	<0.001	0.392	0.585	Supported
SE→AU	0.139	0.410	0.050	2.797	0.005	0.042	0.237	Supported
PV→AU	0.117	0.410	0.056	2.098	0.036	0.008	0.227	Supported

Source: Authors' own compilation.

Note. β : path coefficient; R^2 : explanatory power; SE: standard error; Z: test statistic; CI: confidence interval.

Table 6. Global fit indices and recommended thresholds.

Fit index	Full form	Value	Recommended threshold
$\chi^2(df); p$	Chi-Square Test of Model Fit	$\chi^2 (123) = 376.329$; $p < 0.001$	Non-significant preferred; however, sensitive to sample size – evaluate alongside other indices
CFI	Comparative Fit Index	0.955	≥ 0.95 (good); ≥ 0.90 (acceptable)
TLI	Tucker-Lewis Index	0.944	≥ 0.95 (good); ≥ 0.90 (acceptable)
NNFI	Non-Normed Fit Index (equivalent to TLI)	0.944	≥ 0.95 (good); ≥ 0.90 (acceptable)
RFI	Relative Fit Index	0.918	≥ 0.90 (acceptable)
RMSEA	Root Mean Square Error of Approximation	0.064	≤ 0.06 (good); ≤ 0.08 (acceptable)
NFI	Normed Fit Index	0.934	≥ 0.90 (acceptable)
SRMR	Standardised Root Mean Square Residual	0.035	≤ 0.08 (good); ≤ 0.05 (excellent)
GFI	Goodness-of-Fit Index	0.967	≥ 0.95 (good)

Source: Authors' own compilation.

with standard CB-SEM practice. Figure 2 depicts the measurement and structural models estimated using *lavaan*.

In covariance-based SEM, global fit indices such as the Chi-square test, CFI, TLI, RMSEA, and SRMR assess how well the model-implied covariance matrix reproduces the observed covariance structure, thereby supporting theory confirmation and model adequacy evaluation. In contrast, PLS-SEM does not emphasise global goodness-of-fit, as its primary objective is variance explanation and prediction rather than exact model reproduction. Consequently, model evaluation in *semir* focusses on explained variance (R^2), path coefficient significance, and measurement quality, reflecting its prediction-oriented philosophy.

Semir (PLS-SEM). The PLS-SEM results showed that almost all hypothesised paths were significant (Table 7). Social effects (SE) significantly predicted perceived value (PV), behavioural intention (BI), and actual use (AU), but not attitude (ATT). Perceived benefits (PB) strongly influenced PV, ATT, and BI, while perceived value (PV) was the strongest predictor of attitude and also significantly affected actual

use. Attitude significantly predicted behavioural intention, which in turn strongly predicted actual use. The model demonstrated moderate explanatory power ($R^2 = 0.316\text{--}0.450$), with effect sizes indicating that PV, ATT, and BI were the most influential predictors. All VIF values were well below recommended thresholds, confirming no multicollinearity issues. Unlike *lavaan* (CB-SEM), *semir* does not report global model fit indices, reflecting its prediction-oriented nature. Figure 3 depicts the measurement and structural models estimated using *semir*.

Because PLS-SEM focusses on prediction, the *PLSpredict* (Shmueli et al., 2019) procedure available in *semir* was utilised to evaluate the model's out-of-sample predictive accuracy in addition to its in-sample fit. Tables 8 and 9 present the indicator-level and construct-level predictive results, respectively. The predictive performance of the PLS-SEM model was evaluated using both in-sample and out-of-sample metrics through *PLSpredict*, and the results were compared with a linear regression benchmark (LM). Across all indicators, the PLS model showed only slightly higher RMSE and MAE values in the out-of-sample condition compared to the

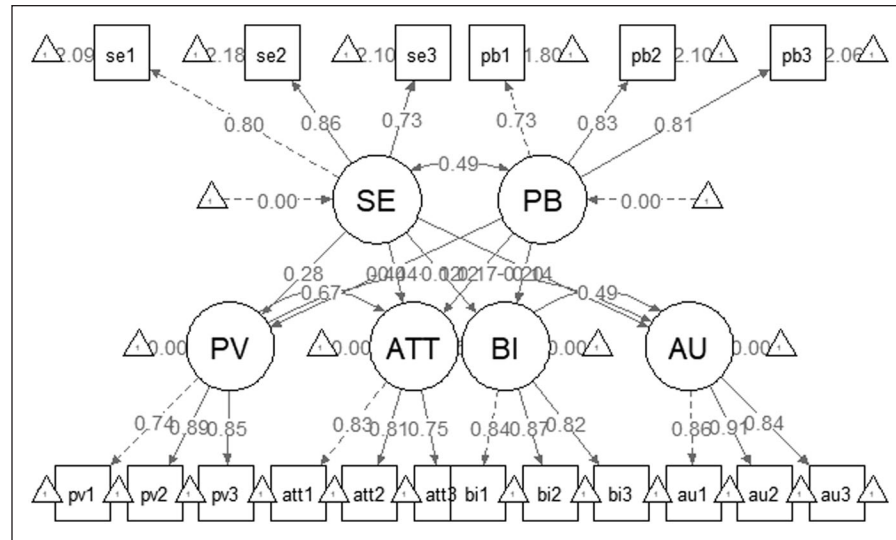


Figure 2. Measurement and structural model results (*lavaan*).
Source: Generated in RStudio using the *semPlot* package (Epskamp, 2025).

Table 7. Hypotheses testing using *Seminr*.

Path	β	R^2	f^2	VIF	B. Mean	B. Std Dev.	T Stat.	2.5% CI	97.5% CI	Decision
SE→PV	0.267	0.316	0.086	1.209	0.267	0.047	5.694	0.173	0.360	Accepted
SE→ATT	0.001	0.450	0.000	1.313	0.002	0.047	0.028	-0.089	0.094	Rejected
SE→BI	0.142	0.424	0.028	1.239	0.142	0.044	3.227	0.056	0.228	Accepted
SE→AU	0.139	0.348	0.023	1.277	0.139	0.046	3.041	0.051	0.228	Accepted
PB→PV	0.396	0.316	0.189	1.209	0.397	0.042	9.439	0.316	0.478	Accepted
PB→ATT	0.176	0.450	0.039	1.437	0.177	0.053	3.305	0.071	0.283	Accepted
PB→BI	0.201	0.424	0.049	1.416	0.200	0.045	4.428	0.110	0.286	Accepted
PV→ATT	0.563	0.450	0.394	1.462	0.564	0.045	12.621	0.475	0.651	Accepted
PV→AU	0.119	0.348	0.015	1.499	0.120	0.048	2.460	0.024	0.214	Accepted
ATT→BI	0.462	0.424	0.284	1.303	0.462	0.045	10.238	0.371	0.547	Accepted
BI→AU	0.446	0.348	0.214	1.415	0.446	0.045	10.003	0.356	0.530	Accepted

Source: Authors' own compilation.

Note. β : path coefficient; R^2 : explanatory power; f^2 : effect size; VIF: variance inflation factor; B.: Bootstrap; Std Dev.: standard deviation; T Stat.: test statistics; CI: confidence interval.

in-sample condition, indicating minimal loss of predictive accuracy when applied to new data. When compared with the LM benchmark, PLS performs competitively, showing marginal differences across most indicators. At the construct level, the differences between in-sample (IS) and out-of-sample (OOS) metrics were very small, with overfitting values ranging from 0.009 to 0.022, which suggests that overfitting is negligible and the model generalises well. Overall, the results support that the PLS-SEM model provides acceptable and stable predictive power for all endogenous constructs (PV, ATT, BI, and AU).

Comparative summary

To enhance methodological transparency, a comparative overview of the CB-SEM model estimated using *lavaan* and the PLS-SEM model estimated using *seminr* was conducted.

Table 10 summarises key methodological, estimation, and reporting characteristics of the two modelling approaches as implemented in the present analysis. Consistent with their established methodological orientations, *lavaan* places greater emphasis on global model fit and theory confirmation, whereas *seminr* focusses on explained variance, prediction-oriented assessment, and flexibility with respect to distributional assumptions.

It is important to note that this distinction reflects default implementation rather than methodological limitation. The *lavaan* package supports robust estimators (eg, MLR and WLSMV) and bootstrapping procedures, allowing CB-SEM models to accommodate non-normal data where required.

Comparative interpretation based on study results. Table 11 reports the coefficients of determination (R^2) for the endogenous constructs obtained from both estimation approaches.

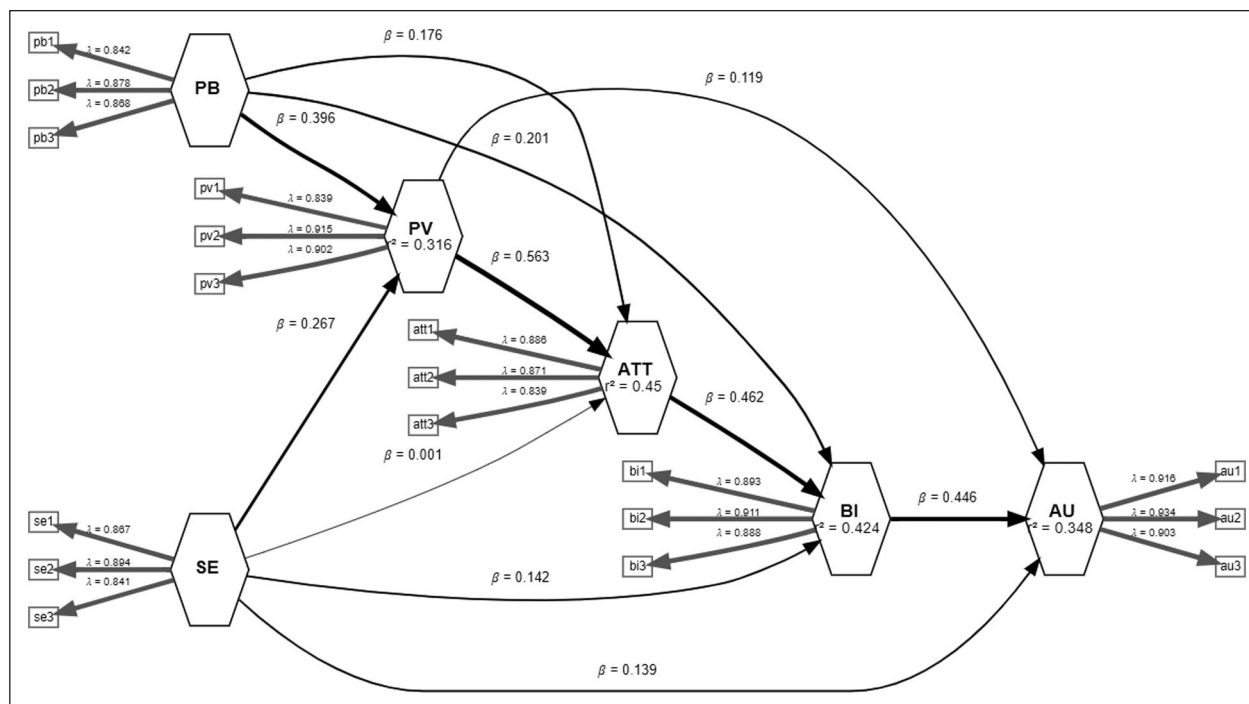


Figure 3. Measurement and structural model results from *seminr*.
Source: Generated in RStudio.

In this illustrative application, the CB-SEM model yielded moderately higher R^2 values than the PLS-SEM model across all endogenous constructs. While PLS-SEM is often characterised as maximising explained variance, prior methodological studies have shown that CB-SEM can produce comparable or even higher R^2 values under conditions of well-specified reflective measurement models and favourable data characteristics (Dash and Paul, 2021; Deng and Yuan, 2023; Rožman et al., 2020; Vuković, 2024).

It is important to emphasise that these differences should be interpreted within the context of the present tutorial demonstration. When indicators exhibit high reliability and the data-generating process aligns closely with the model specification, CB-SEM is capable of efficiently estimating latent covariance structures, which may be reflected in higher explained variance for endogenous constructs (Hair et al., 2021). The observed R^2 patterns therefore serve to illustrate how different estimation philosophies may yield slightly different summaries of model performance when applied to the same model and dataset.

Summary of findings. Overall, both estimation approaches produced substantively consistent results, with similar path directions and comparable levels of statistical significance. The comparison highlights several practical distinctions relevant to applied researchers:

- **Path coefficients:** Estimates were largely comparable in magnitude and direction across both approaches.

- **Explained variance (R^2):** Slightly higher values were observed for *lavaan* in this application, reflecting the characteristics of the data and model specification rather than a general advantage of CB-SEM.
- **Model fit evaluation:** Only *lavaan* provided a comprehensive set of global fit indices, which are central to theory-testing applications.
- **Measurement quality reporting:** *seminr* offered more automated and integrated reporting of reliability and validity metrics, facilitating routine measurement assessment.

Taken together, these results underscore that CB-SEM and PLS-SEM serve complementary methodological purposes. In this illustrative example, CB-SEM was well suited for explanatory model evaluation, whereas PLS-SEM provided a convenient framework for prediction-oriented assessment and detailed measurement diagnostics. The comparison is intended to guide researchers in selecting an appropriate approach based on their research objectives, data characteristics, and reporting priorities, rather than to advocate the superiority of one method over the other.

Discussion and practical implications

The aim of this study was to provide a practical and accessible comparison of two widely used R packages for structural equation modelling – *lavaan* and *seminr* – rather than to benchmark competing estimation paradigms. The paper is

Table 8. Indicator-level predictive performance using PLSpredict.

Construct	Indicator	PLS IS		PLS OOS		LM IS		LM OOS	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
PV	pv1	1.545	1.238	1.553	1.245	1.512	1.227	1.542	1.247
	pv2	1.448	1.162	1.460	1.171	1.442	1.158	1.471	1.181
	pv3	1.534	1.235	1.543	1.242	1.521	1.237	1.545	1.258
ATT	att1	1.547	1.197	1.559	1.205	1.517	1.165	1.558	1.195
	att2	1.435	1.125	1.447	1.132	1.407	1.115	1.455	1.151
	att3	1.465	1.107	1.475	1.113	1.442	1.101	1.475	1.122
BI	bi1	1.478	1.128	1.491	1.138	1.451	1.132	1.491	1.161
	bi2	1.431	1.135	1.442	1.142	1.378	1.091	1.409	1.114
	bi3	1.453	1.111	1.464	1.119	1.433	1.097	1.467	1.120
AU	au1	1.607	1.303	1.625	1.316	1.572	1.251	1.624	1.285
	au2	1.645	1.304	1.664	1.320	1.603	1.271	1.646	1.300
	au3	1.676	1.369	1.692	1.382	1.620	1.311	1.664	1.345

Source: Authors' own compilation.

Note. PLS: partial least squares; IS: in-sample; OOS: out-of-sample; LM: linear model; RMSE: root mean squared error; MAE: mean absolute error.

Table 9. Construct-level predictive performance using PLSpredict.

Metric	PV	ATT	BI	AU
In-sample MSE	0.682	0.548	0.574	0.650
In-sample MAE	0.652	0.563	0.566	0.655
Out-of-sample MSE	0.688	0.558	0.583	0.664
Out-of-sample MAE	0.655	0.567	0.569	0.661
Overfit index	0.009	0.017	0.016	0.022

Source: Authors' own compilation.

Note. MSE: mean squared error; MAE: mean absolute error.

therefore best understood as a reproducible methodological tutorial demonstrating how the same conceptual model can be specified, estimated, and interpreted within different SEM frameworks.

Within this illustrative context, the results highlight the complementary strengths of covariance-based and variance-based approaches. Although both methods yielded substantively consistent conclusions regarding the hypothesised relationships, they differ in emphasis, evaluation logic, and practical implementation. The *lavaan* implementation, grounded in CB-SEM, is particularly suited to confirmatory, theory-driven research where global model fit and theoretical coherence are central. Its comprehensive reporting of fit indices (e.g. CFI, TLI, RMSEA, and SRMR) allows researchers to assess how well a proposed model reproduces the observed covariance structure.

By contrast, *sempr* reflects the prediction-oriented logic of PLS-SEM, prioritising variance explanation, measurement diagnostics, and access to latent variable scores. Its workflow and automated reporting features make it especially attractive in exploratory, applied, and decision-oriented research contexts where predictive relevance is of

primary concern (Hair and Alamer, 2022; Ringle et al., 2023). More broadly, the comparison highlights how differences between the two approaches extend beyond estimation philosophy to practical aspects such as model specification, workflow design, and interpretation of outputs within the R environment.

Although PLS-SEM is often characterised as maximising explained variance (R^2), the present illustrative application produced slightly higher R^2 values for the CB-SEM solution. Under conditions of adequate sample size, strong indicator reliability, and well-specified reflective measurement models, such outcomes are theoretically plausible and have been observed in prior comparative work (Deng and Yuan, 2023; Vuković, 2024). In this tutorial setting, these differences serve to illustrate how data characteristics and estimation logic interact, rather than to support general claims about estimator superiority.

Taken together, the findings emphasise that methodological choices should be guided by research purpose, theoretical maturity, and measurement design rather than methodological preference alone (Hair and Alamer, 2022; Ringle et al., 2023). When applied in alignment with their underlying assumptions, both approaches can provide valuable and complementary insights.

Methodological contribution

This study contributes to the SEM literature by offering a fully reproducible, side-by-side comparison of *lavaan* and *sempr* within a single open-source environment (R). By documenting data generation, model specification, estimation steps, and software versions, the paper lowers the barrier for applied researchers, particularly those new to SEM, who seek practical guidance in selecting and implementing appropriate modelling tools.

Table 10. Comparison of CB-SEM (*lavaan*) and PLS-SEM (*semnr*).

Aspect	<i>lavaan</i> (CB-SEM)	<i>semnr</i> (PLS-SEM)
SEM type	Covariance-based Structural Equation Modelling (CB-SEM)	Variance-based Partial Least Squares Structural Equation Modelling (PLS-SEM)
Sample size suitability	Requires larger samples; assumes multivariate normality (eg, $N > 200$)	Performs well with small to moderate samples; robust to non-normal data
Syntax	Formula-like and compact	Verbose but intuitive and user-friendly
Model fit indices	Provides multiple global fit indices (χ^2 , CFI, TLI, RMSEA, SRMR)	No global fit indices; focusses on predictive measures (R^2 and PLSpredict)
Validity testing	Does not automatically report AVE, CR, or discriminant validity; can be computed via <i>semTools</i>	Automatically reports AVE, Composite Reliability, Cronbach's α , Fornell-Larcker, and HTMT
Multicollinearity testing	Does not directly report VIF; must be computed manually	Reports VIF for indicators and constructs
Primary objective	Theory testing and model confirmation	Prediction and theory development
Bootstrapping	Supported for standard errors and confidence intervals	Widely used for significance testing of paths and loadings
Latent score estimation	Not emphasised; latent scores can be derived post hoc	Directly estimated as part of the PLS algorithm
Estimation method	Maximum Likelihood (assumes normality)	Partial Least Squares algorithm (nonparametric)
Results emphasis	Detailed parameter estimates and model-fit diagnostics	Emphasises path coefficients, R^2 , and predictive power

Source: Authors' own compilation.

Table 11. Explained variance (R^2) comparison across estimation methods.

Endogenous construct	<i>semnr</i> (PLS-SEM)	<i>lavaan</i> (CB-SEM)
PV	0.316	0.393
ATT	0.450	0.578
BI	0.424	0.525
AU	0.348	0.410

Source: Authors' own compilation.

Rather than advancing new methodological claims, the contribution lies in translating conceptual distinctions between CB-SEM and PLS-SEM into transparent and replicable analytical workflows. In doing so, the study moves beyond abstract methodological comparisons by demonstrating how these differences manifest in practice, complementing prior discussions in the methodological literature (Dash and Paul, 2021; Vuković, 2024).

Limitations and future directions

Several limitations should be acknowledged. First, the analysis relies on synthetic data with reflective indicators and a relatively simple model structure, chosen to maximise tutorial clarity rather than methodological generalisation. Second, the study does not constitute a formal simulation design comparing estimator performance under varying conditions.

Future research could extend this comparison by incorporating formative constructs, nonlinear relationships, multi-group analyses, and systematic variations in sample size or

distributional assumptions. Applying the same comparative framework to empirical datasets would further assess the generalisability of the illustrative differences observed here and build upon existing comparative investigations (Dash and Paul, 2021; Deng and Yuan, 2023; Vuković, 2024).

Conclusion

This paper set out to provide a clear, reproducible, and practically oriented comparison of two widely used R packages for structural equation modelling – *lavaan* (CB-SEM) and *semnr* (PLS-SEM) – by specifying and estimating the same conceptual model using both approaches. Rather than advocating for a particular modelling paradigm, the study illustrates how similar theoretical models can be operationalised within different SEM frameworks, each grounded in distinct methodological philosophies and analytical priorities.

The results demonstrate that, while both approaches yielded substantively consistent conclusions regarding the hypothesised relationships, they differ meaningfully in terms of estimation logic, model evaluation, and reporting conventions. These differences are reflected not only in their underlying assumptions but also in how models are specified, estimated, and interpreted in practice within the R environment. In particular, *lavaan* provides a covariance-based framework with comprehensive global fit assessment, while *semnr* emphasises variance explanation, measurement diagnostics, and accessible workflows for applied research.

Importantly, the comparison does not suggest that one approach is universally superior to the other. Instead, it highlights the importance of aligning methodological choices with the primary research goal, data characteristics, and

measurement design. By offering a hands-on, side-by-side demonstration within a single open-source environment, this study provides applied researchers, particularly those working in R, with practical guidance for making informed and transparent methodological decisions.

It is also important to note that the present study does not constitute a simulation-based evaluation of estimator performance, nor does it aim to establish general claims about the relative strengths of CB-SEM and PLS-SEM. The use of synthetic data serves a pedagogical purpose, enabling the modelling steps and software behaviour to be demonstrated in a controlled and reproducible manner. Future research could build on this tutorial foundation by incorporating systematic simulation designs, applying the comparison to diverse empirical datasets, or integrating predictive validation and hybrid approaches that combine SEM with machine learning techniques.

Overall, this paper aims to serve as a practical entry point for researchers seeking to understand and implement SEM in R, and to support more informed, transparent, and methodologically appropriate use of SEM tools in applied research.

Acknowledgements

The authors acknowledge the use of ChatGPT-5.2 for copy-editing and language refinement in specific parts of the manuscript. All conceptual, analytical, and interpretive components were solely developed by the authors.

ORCID iDs

Isaac Osei  <https://orcid.org/0009-0008-0159-4206>

Dennis Opoku Boadu  <https://orcid.org/0009-0005-8220-4086>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

The data analysed in this study can be obtained from the corresponding author* upon reasonable request.

References

- Almeida F (2024) Editorial: Performing a structural equation modeling (SEM) in innovation science studies. *International Journal of Innovation Science* 16(6): 1005–1011. <https://doi.org/10.1108/IJIS-12-2024-289>
- Bollen KA (1989) *Structural Equations with Latent Variables*. Wiley.
- Bollen KA and Diamantopoulos A (2017) In defense of causal-formative indicators: A minority report. *Psychological Methods* 22(3): 581–596. <https://doi.org/10.1037/met0000056>
- Cepeda G, Roldán JL, Sabol M, et al. (2024) Emerging opportunities for information systems researchers to expand their PLS-SEM analytical toolbox. *Industrial Management and Data Systems* 124(6): 2230–2250. <https://doi.org/10.1108/IMDS-08-2023-0580>
- Chin WW (1998) The partial least squares approach to structural equation modelling. In: Marcoulides GA (ed.) *Modern Methods for Business Research*. Lawrence Erlbaum Associates, pp.295–336.
- Dash G and Paul J (2021) CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change* 173: 121092. <https://doi.org/10.1016/j.techfore.2021.121092>
- Deng L and Yuan K-H (2023) Which method is more powerful in testing the relationship of theoretical constructs? A meta comparison of structural equation modeling and path analysis with weighted composites. *Behavior Research Methods* 55(3): 1460–1479. <https://doi.org/10.3758/s13428-022-01838-z>
- Epskamp S (2025) *semPlot: Path Diagrams and Visual Analysis of Various SEM Packages' Output*. CRAN. <https://doi.org/10.32614/CRAN.package.semPlot>
- Fornell C and Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18(1): 39–50. <https://doi.org/10.2307/3151312>
- Goktas P and Dirsehan T (2025) Using PLS-SEM and XAI for causal-predictive services marketing research. *Journal of Services Marketing* 39(1): 53–68. <https://doi.org/10.1108/JSM-10-2023-0377>
- Gudergan SP, Moisescu OI, Radomir L, et al. (2025) Special issue editorial: Advanced partial least squares structural equation modeling (PLS-SEM) applications in business research. *Journal of Business Research* 188: 115087. <https://doi.org/10.1016/j.jbusres.2024.115087>
- Guenther P, Guenther M, Ringle CM, et al. (2023) Improving PLS-SEM use for business marketing research. *Industrial Management and Data Systems* 111: 127–142. <https://doi.org/10.1016/j.indmarman.2023.03.010>
- Hair J and 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. <https://doi.org/10.1016/j.rmal.2022.100027>
- Hair JF, Babin BJ and Krey N (2017) Covariance-based structural equation modeling in the journal of advertising: Review and recommendations. *Journal of Advertising* 46(1): 163–177. <https://doi.org/10.1080/00913367.2017.1281777>
- Hair JF, Babin BJ, Ringle CM, et al. (2025) Covariance-based structural equation modeling (CB-SEM): A SmartPLS 4 software tutorial. *Journal of Marketing Analytics* 13: 709–724. <https://doi.org/10.1057/s41270-025-00414-6>
- Hair JF, Hult GTM, Ringle CM, et al. (2021) *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-80519-7>
- Hair JF, Hult GTM, Ringle CM, et al. (2022) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed. Sage Publications.
- Hair JF, Risher JJ, Sarstedt M, et al. (2019) When to use and how to report the results of PLS-SEM. *European Business Review* 31(1): 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hair JF, Jr, Sarstedt M, Hopkins L, et al. (2014) Partial least squares structural equation modeling (PLS-SEM): An emerging tool in

- business research. *European Business Review* 26(2): 106–121. <https://doi.org/10.1108/EBR-10-2013-0128>
- Henseler J, Ringle CM and Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science* 43(1): 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu L and Bentler P (1998) Fit indices sensitivity to misspecification. *Psychological Methods* 3(4): 424–453.
- Jorgensen TD, Pornprasertmanit S, Schoemann AM, et al. (2025) SemTools: Useful tools for structural equation modeling. R package version 0.5-7. Available at: <https://CRAN.R-project.org/package=semTools> (accessed 27 September 2025).
- Kline RB (2016) *Principles and Practice of Structural Equation Modeling*, 4th edn. The Guilford Press.
- Kline RB (2023) *Principles and Practice of Structural Equation Modeling*, 5th edn. The Guilford Press.
- Lohmöller J-B (1989) *Latent Variable Path Modeling with Partial Least Squares*. Physica-Verlag HD. <https://doi.org/10.1007/978-3-642-52512-4>
- McDonald RP (1996) Path analysis with composite variables. *Multivariate Behavioral Research* 31(2): 239–270. https://doi.org/10.1207/s15327906mbr3102_5
- R Core Team (2025) R: A Language and Environment for Statistical Computing. Available at: <https://www.r-project.org/> (accessed 9 July 2025).
- Ray S, Danks NP and André CV (2021) *SEMinR: Domain-specific Language for Building, Estimating, and Visualizing Structural Equation Models in R*. SSRN, pp.1–16. <https://doi.org/doi.org/10.2139/ssrn.3900621>
- Ray S, Danks NP and André CV (2025) *seminr: Building and Estimating Structural Equation Models* (R package version 2.3.4). Available at: <https://cran.r-project.org/package=seminr> (accessed 27 September 2025).
- Reinartz W, Haenlein M and Henseler J (2009) An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing* 26(4): 332–344. <https://doi.org/10.1016/j.ijresmar.2009.08.001>
- Richter NF and Tudoran AA (2024) Elevating theoretical insight and predictive accuracy in business research: Combining PLS-SEM and selected machine learning algorithms. *Journal of Business Research* 173: 114453. <https://doi.org/10.1016/j.jbusres.2023.114453>
- Rigdon EE, Sarstedt M and Ringle CM (2017) On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing ZFP* 39(3): 4–16. <https://doi.org/10.15358/0344-1369-2017-3-4>
- Ringle CM, Sarstedt M, Sinkovics N, et al. (2023) A perspective on using partial least squares structural equation modelling in data articles. *Data in Brief* 48: 109074. <https://doi.org/10.1016/j.dib.2023.109074>
- Rosseel Y (2012) lavaan : An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel Y, Jorgensen TD and De Wilde L (2025) lavaan: Latent variable analysis. *CRAN: Contributed Packages* (R package version 0.6-20). <https://doi.org/10.32614/CRAN.package.lavaan>
- Rožman M, Tominc P and Milfelner B (2020) A comparative study using two SEM techniques on different samples sizes for determining factors of older employee's motivation and satisfaction. *Sustainability* 12(6): 2189. <https://doi.org/10.3390/su12062189>
- Sakaria D, Maat SM and Mohd Matore MEE (2023) Examining the optimal choice of SEM statistical software packages for sustainable mathematics education: A systematic review. *Sustainability* 15(4): 3209. <https://doi.org/10.3390/su15043209>
- Sarstedt M, Adler SJ, Ringle CM, et al. (2024) Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of Product Innovation Management* 41(6): 1100–1117. <https://doi.org/10.1111/jpim.12738>
- Sarstedt M, Hair JF, Pick M, et al. (2022) Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology and Marketing* 39(5): 1035–1064. <https://doi.org/10.1002/mar.21640>
- Sarstedt M, Hair JF, Ringle CM, et al. (2016) Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research* 69(10): 3998–4010. <https://doi.org/10.1016/j.jbusres.2016.06.007>
- Sarstedt M, Ringle CM and Hair JF (2017) Partial least squares structural equation modeling. In: Homburg C, Klarmann M and Vomberg A (eds) *Handbook of Market Research*. Springer International Publishing, pp.1–40. https://doi.org/10.1007/978-3-319-05542-8_15-1
- Schuberth F, Rosseel Y, Rönkkö M, et al. (2023) Structural parameters under partial least squares and covariance-based structural equation modeling: A comment on Yuan and Deng (2021). *Structural Equation Modeling* 30(3): 339–345. <https://doi.org/10.1080/10705511.2022.2134140>
- Schuberth F, Schamberger T, Kemény I, et al. (2025) The sum score model: Specifying and testing equally weighted composites using structural equation modeling. *Psychometrika* 90(1): 358–383. <https://doi.org/10.1017/psy.2024.5>
- Sharma PN, Sarstedt M, Ringle CM, et al. (2024) A framework for enhancing the replicability of behavioral MIS research using prediction oriented techniques. *International Journal of Information Management* 78: 102805. <https://doi.org/10.1016/j.ijinfomgt.2024.102805>
- Shmueli G, Sarstedt M, Hair JF, et al. (2019) Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing* 53(11): 2322–2347. <https://doi.org/10.1108/EJM-02-2019-0189>
- Spearman C (1904) “General intelligence,” objectively determined and measured. *The American Journal of Psychology* 15(2): 201–292. <https://doi.org/doi.org/10.2307/1412107>
- Vinzi VE, Chin WW, Henseler J, et al. (eds) (2010) *Handbook of Partial Least Squares*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-32827-8>
- Vuković M (2024) CB-SEM vs PLS-SEM comparison in estimating the predictors of investment intention. *Croatian Operational Research Review* 15(2): 131–144. <https://doi.org/10.17535/corr.2024.0011>
- Wold H (1975) Path models with latent variables: The NIPALS approach. In *Quantitative Sociology*. Elsevier, pp.307–357. <https://doi.org/10.1016/B978-0-12-103950-9.50017-4>
- Wright S (1921) Correlation and causation. *Journal of Agricultural Research* 20: 557–585.

Author biographies

Isaac Osei is a Computer and Data Scientist with more than 12 years of experience in the IT industry and academia. He earned his bachelor's degree in Computer Science from the University of Ghana, Legon, and later completed a Master of Computer Applications at Amity University, Noida, India. He is currently pursuing his PhD at SRM University-AP, India. His research interests span the application of Structural Equation Modelling (SEM) in Human-Computer Interaction (HCI) and Financial Technology, as well as Machine Learning, Deep Learning, Bioinformatics, and Applied Statistics.

Chettupally Anil Carie is a Computer Scientist and academic with extensive experience in teaching and research. He earned his bachelor's degree in computer science and engineering from Ramappa Engineering College, JNTU Hyderabad, and went on to complete his Master of Technology at Karunya University of Technology, Coimbatore. He later received his PhD from Dalian University of Technology, China. He is currently an Associate Professor in the Department of Computer Science and Engineering at SRM University-AP, India. His research interests span Application of Structural Equation Modelling in Information Systems, Intelligent Internet of Things (IoT), vehicular and smart-city IoT systems, Deep Learning, advanced routing and resource-allocation algorithms, and network optimization for socially impactful applications.

Satish Anamalamudi received the B.Eng. degree in Computer Science and Engineering from JNTU Hyderabad, the M.Tech. degree from Karunya University, and the Ph.D. degree in Communication and Information Systems from the Dalian University of Technology, China. He has professional experience as a Research Engineer at Huawei Technologies, Beijing. Currently, he is an Associate Professor at SRM University-AP, India, where he also serves as Director (I/C) – ITKM. His research focuses on MAC and routing protocols for cognitive radio ad hoc networks, IoT, and 5G systems.

Dennis Opoku Boadu is a Computer Science Lecturer and Researcher at the University of Ghana and Ghana Christian University College. He holds an MPhil in Computer Science from the Kwame Nkrumah University of Science and Technology, an M.A in Development Communication at Ghana Institute of Journalism, a BSc in Computer Science at the University of Ghana, Legon and also a PhD Candidate in Computer Science at the University of Ghana. His research spans machine learning, graph-based computational models, and virtual reality applications in the fields of health and education. He explores how AI, bioinformatics, and blockchain security can drive innovation and social transformation. Dennis has published and reviewed several works in journals such as Operations Research Forum, IGI Global, and PLOS. Passionate about mentoring and system design, he combines academic rigour with practical implementation to bridge data science, technology, and human development.

Michael Kwakye is an accomplished IT professional and academic with over ten years of experience in higher education. He's a Doctoral Candidate with Masters in IT Application Development from University of Sunderland. He currently serves as the Head of the IT Department at Ghana Christian University, where he oversees strategic technology operations, digital transformation initiatives, and the implementation of innovative IT solutions to enhance teaching, learning, and administrative efficiency. Alongside his administrative role, he is also a dedicated Lecturer, committed to training the next generation of computing professionals through engaging and practical instruction. His expertise spans several domains including Software Usability Engineering, Database Administration and Statistics, with strong technical proficiency in Structural Equation Modelling, R, Java, and PHP. Over the years, he has contributed to multiple academic and system development projects, demonstrating a deep understanding of user-centered design and software development principles. Beyond his professional work, he has a passion for reading and writing articles, often reflecting on technology trends, usability challenges, and the future of computing in education.

Appendix A: Synthetic data generation using *lavaan*

Table A1. Synthetic data generation for SEM demonstration.

```

# Load required package
library(lavaan)

# Set random seed for full reproducibility
set.seed(123)

# -----
# Population model specification
# -----
# Measurement model: six reflective latent constructs
# Structural model: sequential relationships among constructs
population_model <- '
  # Measurement model
  SE =~ 0.85*se1 + 0.82*se2 + 0.80*se3
  PB =~ 0.85*pb1 + 0.83*pb2 + 0.80*pb3
  PV =~ 0.85*pv1 + 0.88*pv2 + 0.85*pv3
  ATT =~ 0.88*att1 + 0.85*att2 + 0.82*att3
  BI =~ 0.88*bi1 + 0.90*bi2 + 0.88*bi3
  AU =~ 0.90*au1 + 0.92*au2 + 0.90*au3

  # Structural model
  PB ~ 0.40*SE
  PV ~ 0.45*PB
  ATT ~ 0.50*PV
  BI ~ 0.55*ATT
  AU ~ 0.60*BI
'

# -----
# Simulate continuous data
# -----
synthetic_data_continuous <- simulateData(
  model = population_model,
  sample.nobs = 500,
  meanstructure = TRUE
)

# -----
# Discretise indicators to 7-point Likert-type scale
# -----
# This step reflects common practice in applied SEM research.
# Values are rounded and constrained to the 1-7 range.
synthetic_data_likert <- as.data.frame(
  apply(synthetic_data_continuous, 2, function(x) {
    x_round <- round(x)
    x_round[x_round < 1] <- 1
    x_round[x_round > 7] <- 7
    return(x_round)
  })
)

# -----
# Add respondent ID
# -----
synthetic_data_likert$ID <- seq_len(nrow(synthetic_data_likert))
# Reorder columns to place ID first
synthetic_data_likert <- synthetic_data_likert[,
  c("ID", setdiff(names(synthetic_data_likert), "ID"))
]

# -----
# Preview final dataset
# -----

df <- synthetic_data_likert

head(df)

# The resulting dataset is used for all analyses reported
# in the manuscript using lavaan (CB-SEM) and semr (PLS-SEM).

```
