A COMPARISON BETWEEN COVARIANCE - BASED (CB) AND PARTIAL LEAST SQUARES (PLS) STRUCTURAL EQUATION MODELING : APPLICATIONS, ASSUMPTIONS, AND APPROPRIATENESS

Sipnarong Kanchanawongpaisan *[1]
*[1] Suan Sunandha Rajabhat University
* Email: sipnarong.ka@gmail.com

Received 4 August 2024 *Revised 29 August 2024* *Accepted 6 September 2024*

Abstract

This study compares Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM). We emphasize the distinct scopes, techniques, and applications of each. CB-SEM is specifically designed to evaluate hypotheses and validate them. It is effective with substantial sample sizes and data sets commonly disseminated and extensively used chiefly in social sciences, psychology, and education. The maximum likelihood estimation approach estimates parameters, emphasizing fit indices such as Chi-square, RMSEA, CFI, and TLI. In contrast, Partial Least Squares Structural Equation Modeling (PLS-SEM) is particularly suitable for doing exploratory research or making predictions. It can handle lower sample sizes and data that does not follow a normal distribution. Furthermore, it is prevalent in business, marketing science, information systems, and management research. An iterative technique optimizes the explained variance in Partial Least Squares Structural Equation Modeling (PLS-SEM). This algorithm focuses on key features such as R-squared, Q-squared, Average Variance Extracted (AVE), and Composite Reliability (CR), which are indicators of prediction quality. This research examines how measurement error is addressed in CB-SEM by explicitly including terms, whereas PLS-SEM takes an implicit method. Furthermore, it examines model fit indices and data needs to assist researchers in selecting between these two methodologies, providing them with a practical reference. This document offers comprehensive guidance that considers the study goals and the specific features of the accessible data sets. The preceding discourse enhances the use of rigorous techniques and offers valuable direction to researchers.

Keywords : Structural Equation Modeling (SEM), Covariance-Based SEM (CB-SEM), Partial Least Squares SEM (PLS-SEM)

Introduction

SEM is a comprehensive multivariate technique that analyzes complicated relationships among observed and latent variables (Little, 2023). It is not just its ability to scrutinize and quantify links between numerous variables that have made it indispensable in social sciences, business research, psychology, and education; it has also made possible the combination of direct and indirect effects. Such trait that makes SEM different from other statistical techniques allows for a better perception of the theoretical model's underpinning structure (Sharma et al., 2023). Apart from sharing common principles, the two main approaches to SEM (CB-SEM) Covariance-Based SEM are distinct concerning their target objectives, estimation methods, and application contexts. While there are many similarities, CB-SEM is mainly based on reproducing the covariance matrix of observed data and can be considered theory testing or confirmation-oriented. In contrast, PLS-SEM seeks to maximize explained variance in dependent variables with an orientation towards theory development and prediction (Henseler & Schuberth, 2024).

Despite the widespread use of both CB-SEM and PLS-SEM, a significant study gap urgently needs to be addressed. There is an urgent need to fill a significant knowledge gap. For example, existing studies are not comprehensive or comparative; hence, they do not provide the context necessary for making judgments based on the appropriateness of either method under different conditions. Nevertheless, choosing between them can be very difficult, especially for researchers who are beginners in SEM (Structural Equation Modelling). Failing to choose the most appropriate method between these two alternatives is further complicated by the literature that has made varying recommendations and contradicted themselves most of the time, necessitating our investigation (Sabole et al.,2023). These studies mainly focus on its technical aspects but do not sufficiently address some practical implications arising from one's preference. This further implies that we need to offer tangible guidance when deciding which method to prefer between CB-SEM and PLS-SEM, given the increasing complexity of research models and adoption by new fields. The study we conduct aims to fill this void by providing researchers with adequate tools to guide them into making more informed choices and reducing confusion, hence improving the rigor of research methods employed within our field. By so doing, there will be a significant improvement in research methodology within our field, resulting in more robust findings.

This study aims to give practical direction to researchers by comparing CB-SEM and PLS-SEM in detail to inform them which SEM technique can best suit them based on their objectives, nature of data, and theoretical basis. In addition, there are many misconceptions about what CB-SEM and PLS-SEM can or cannot do. This paper will explain some of these misconceptions by explaining the two methods' strengths and limitations (Cepeda et al., 2024). This study will plug a gap in comparative literature and add to a better understanding of the reasons why or when one should opt for CB-SEM vs. PLS-SEM. Methodological rigor across diverse research fields requires such knowledge. When research models get more complicated, it is necessary to choose appropriate analytical techniques that will produce valid and reliable results. This study will improve methodological rigor by assisting researchers in making appropriate choices regarding their analytic strategy. Moreover, technical literature provides detailed discussions on CB-SEM versus PLS-SEM along specialized lines. A broader perspective will make this work accessible to all readers with different areas of specialization within various fields of science (Hair et al., 2019).

Conceptual Differences Between CB-SEM and PLS-SEM

Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM) are sturdy strategies for analyzing complicated relationships among determined and latent variables. However, they fluctuate fundamentally in their conceptual method, goals, and underlying philosophy. Understanding those variations is essential for selecting the best technique for a given research context.

1. Objective and Philosophy

CB-SEM is rooted in classical statistical theory and is designed primarily for theory testing and confirmation. Its main objective is to reproduce the covariance matrix of the observed data as closely as possible. This approach is inherently confirmatory, testing whether the data fits a predefined theoretical model. CB-SEM aims to validate the proposed relationships between variables based on solid theoretical foundations by focusing on reproducing the covariance structure. It uses various goodness-of-fit indices to determine how well the model represents the data, thus ensuring that the theoretical constructs are supported by empirical evidence (Hair et al., 2014). In contrast, PLS-SEM is oriented towards prediction and theory development. Its primary objective is to maximize the explained variance of the dependent variables. PLS-SEM is inherently exploratory, identifying relationships and patterns within the data that can inform theory building. This method does not aim to replicate the observed covariance matrix but instead seeks to explain the variances of the endogenous constructs. PLS-SEM is particularly useful when the theoretical framework is not well-established, and the researcher aims to explore or refine new models. It provides flexibility in model specification and can handle complex models with many indicators and constructs, making it a powerful tool for predictive modeling and exploratory research (Sarstedt et al., 2024).

1. Model Fit

CB-SEM places a strong emphasis on the overall fit of the model to the observed data. It provides a comprehensive set of fit indices to assess how well the hypothesized model replicates the covariance matrix of the observed variables.

Chi-square (χ^2): This statistic tests the null hypothesis that the model fits the data perfectly. A significant Chi-square value indicates a poor fit. However, this test is sensitive to sample size; with large samples, even minor discrepancies can result in significant Chi-square values (Kline, 2015).

$$\chi^2 = (N - 1) \times F$$

where N is the sample size, and F is the minimum value of the discrepancy function that measures the difference between the observed and model-implied covariance matrices. A significant Chi-square value indicates a poor fit. However, this test is sensitive to sample size; with large samples, even minor discrepancies can result in significant Chi-square values.

Root Mean Square Error of Approximation (RMSEA): RMSEA measures the extent to which the model, with unknown but optimally chosen parameter estimates, fits the population covariance matrix. Values less than 0.08 indicate a reasonable fit, and values below 0.05 indicate a good fit (Browne & Cudeck, 1993).

$$\text{RMSEA} = \sqrt{\frac{\chi^2/df - 1}{N - 1}}$$

Where df is the degree of freedom, values less than 0.08 indicate a reasonable fit, and values below 0.05 indicate a good fit.

Comparative Fit Index (CFI): CFI compares the fit of the user-specified model to a more restricted, nested baseline model. Values close to 1.0 indicate a good fit, with values above 0.90 generally considered acceptable (Bentler, 1990).

$$\text{CFI} = 1 - \frac{(\chi^2_{\text{model}} - df_{\text{model}})}{(\chi^2_{\text{baseline}} - df_{\text{baseline}})}$$

Values close to 1.0 indicate a good fit, with values above 0.90 generally considered acceptable.

Tucker-Lewis Index (TLI): Similar to CFI, TLI adjusts for model complexity, penalizing for adding parameters that do not significantly improve the model fit. Values above 0.90 typically indicate a good fit (Tucker & Lewis, 1973).

$$\text{TLI} = \frac{\chi^2_{\text{baseline}}/df_{\text{baseline}} - \chi^2_{\text{model}}/df_{\text{model}}}{\chi^2_{\text{baseline}}/df_{\text{baseline}} - 1}$$

Values above 0.90 are typically considered indicative of a good fit.

These indices in CB-SEM collectively help researchers determine if their theoretical model is a plausible representation of the data structure, ensuring that the hypothesized relationships among variables are supported by empirical evidence.

In contrast, PLS-SEM does not prioritize overall model fit similarly. Instead, it focuses on the model's predictive power, emphasizing measures that indicate how well the model explains the variance in the dependent variables. Key indicators in PLS-SEM include as follows:

R-squared (R^2): This measure indicates the proportion of variance in the endogenous constructs explained by the exogenous constructs. Higher R^2 values suggest better explanatory power (Hair et al., 2017).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the observed data. Higher R^2 values suggest better explanatory power.

Cross-validated Redundancy (Q^2): Q^2 uses a blindfolding procedure to assess the model's predictive accuracy. Positive Q^2 values indicate the model is predictively relevant for a given endogenous construct (Chin, 1998).

$$Q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Positive Q^2 values indicate the model has predictive relevance for a given endogenous construct.

Average Variance Extracted (AVE): AVE measures the variance a construct captures concerning the amount of variance due to measurement error.

Values above 0.50 are considered acceptable, indicating that the construct explains more than half of the variance of its indicators (Hair et al., 2010).

$$\mathrm{AVE} = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum \theta_i}$$

Where λ_i are the factor loadings and θ_i are the error variances. Values above 0.50 are considered acceptable, indicating that the construct explains more than half of the variance of its indicators.

Composite Reliability (CR): CR evaluates the internal consistency of the constructs, similar to Cronbach's alpha. Values above 0.70 are generally considered satisfactory (Hair et al., 2017).

$$\mathrm{CR} = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_i}$$

Values above 0.70 are generally considered satisfactory.

PLS-SEM's focus on these measures reflects its primary goal of maximizing explained variance and predictive accuracy rather than fitting a predefined covariance structure. This makes PLS-SEM particularly useful in exploratory research where the theoretical model is still being developed and refined.

Aspect	CB-SEM	PLS-SEM
Model Fit Emphasis	The overall fit of the model to the data	Predictive power and explained variance
Key Fit Indices	Chi-square (χ^2), RMSEA, CFI, TLI	R-squared (R^2), Q-squared (Q^2), AVE, CR
Approach to Fit	Confirmatory, goodness-of-fit indices	Exploratory, predictive relevance measures

Table 1 *Summary of Model Fit Differences*

2. Estimation Method

Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM) use separate estimate approaches that align with their respective aims and philosophies. Comprehending these techniques is essential for choosing a suitable methodology for a particular study setting. CB-SEM often uses Maximum Likelihood Estimation (MLE) to estimate model parameters. The Maximum probability Estimation (MLE) method seeks to identify the parameter values that optimize the probability of witnessing the sample data, given the model. This approach relies on many fundamental assumptions and scientific procedures.

The objective of MLE

The primary objective of MLE is to identify parameter estimates that make the observed data most probable under the specified model. This involves calculating the likelihood function, which represents the probability of the observed data as a function of the model parameters (Bollen, 1989).

Likelihood Function

The likelihood functions $L(\theta)$ for a set of parameters θ is given by:

$$L(\theta) = \prod_{i=1}^{N} f(x_i|\theta)$$

Where $f(x_i|\theta)$ is the probability density function of the observed data x_i given the parameters θ, and N is the sample size.

Maximizing the Likelihood

To find the parameter estimates that maximize the likelihood function, MLE solves the following optimization problem.

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

This involves taking the natural logarithm of the likelihood function (to simplify calculations) and finding the parameter values that maximize this log-likelihood

$$\log L(\theta) = \sum_{i=1}^{N} \log f(x_i|\theta)$$

Assumptions of MLE

Normality: MLE assumes that the data are normally distributed. This assumption is critical because the properties of the maximum likelihood estimators, such as consistency, efficiency, and asymptotic normality, rely on it (Kline, 2015).

Large Sample Size: MLE requires a large sample size to produce reliable estimates. This is because the asymptotic properties of MLE ensure that the estimators are unbiased and have minimum variance in large samples, only holding when the sample size is sufficiently large (Jöreskog, 1970)

Challenges and Limitations

Sensitivity to Sample Size: Based on MLE, the Chi-square statistic used in CB-SEM is sensitive to sample size. Even minor discrepancies between the model and the observed data can result in a significant Chi-square value in large samples, indicating a poor fit (Hu & Bentler, 1999).

Assumption Violations: When the normality assumption is violated, MLE can produce biased parameter estimates and incorrect standard errors, leading to invalid conclusions. Alternative estimation methods, such as robust maximum likelihood or bootstrapping, can address this issue (Byrne, 2016).

Aspect	CB-SEM	PLS-SEM
Estimation Method	Maximum Likelihood Estimation (MLE)	Iterative algorithm combining PCA and OLS regression
Assumptions	Normality, large sample size	No assumption of normality, suitable for smaller samples
Objective	Maximizing the likelihood of observed data	Maximizing explained variance and predictive accuracy

Table 2 *Summary of Estimation Method Differences*

Therefore, CB-SEM's use of MLE is geared towards achieving the most probable parameter estimates given the observed data, assuming normality and large sample sizes. This method ensures the accuracy and reliability of the model's parameters but also imposes strict requirements on the data.

3. Data Requirements

Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM) have specific data needs corresponding to their unique aims and estimate methodologies. Comprehending these criteria is essential for researchers to choose the suitable methodology for their study environment. CB-SEM and PLS-SEM have distinct differences in their assumptions on data properties, namely in sample size and data distribution.

CB-SEM Data Requirements

Sample Size:

CB-SEM typically requires a large sample size to produce reliable and valid results. This is because the accuracy of parameter estimates, the power of statistical tests, and the stability of model fit indices improve with larger samples. A general rule of thumb is to have at least 200 cases, although more complex models may require larger samples (Kline, 2015).

Minimum Sample Size: CB-SEM's minimum recommended sample size is often cited as 200. However, some authors suggest a minimum of 10 to 20 times the model's estimated parameters (Hair et al., 2010).

Normality Assumptions:

CB-SEM assumes that the data are multivariate normally distributed. This assumption is critical because the maximum likelihood estimation (MLE) method used in CB-SEM relies on the normality of the data to produce unbiased and efficient parameter estimates (Byrne, 2016).

Multivariate Normality: The assumption of multivariate normality means that each observed variable should be normally distributed, and any linear combination of the variables should also be distributed.

Implications of Non-Normality: When data deviate from normality, parameter estimates and standard errors may be biased, leading to incorrect conclusions. Researchers may use alternative estimation methods, such as robust maximum likelihood (RML) or bootstrapping, to address non-normality (Kline, 2015).

Missing Data:

CB-SEM also assumes that data are missing at random (MAR) or entirely at random (MCAR). If this assumption is violated, the results can be biased. Various imputation techniques or the complete information maximum likelihood (FIML) method can handle missing data appropriately (Little & Rubin, 2019).

PLS-SEM Data Requirements

Sample Size:

PLS-SEM is more flexible regarding sample size requirements and can produce reliable results with smaller samples. This flexibility makes PLS-SEM particularly useful in exploratory research and studies with limited sample sizes (Hair et al., 2017).

Minimum Sample Size: A standard guideline for PLS-SEM is the "10-times rule," which suggests that the minimum sample size should be ten times the maximum number of paths pointing to any construct in the model. However, recent research indicates that sample sizes of 100 to 200 are generally sufficient for most PLS-SEM applications (Hair et al., 2017).

Normality Assumptions:

PLS-SEM does not assume that the data are typically distributed. This lack of reliance on normality makes PLS-SEM robust to deviations from normality and suitable for analyzing non-normal, skewed, or kurtotic data (Henseler et al., 2015).

Robustness to Non-Normality: PLS-SEM's iterative algorithm, which combines principal component analysis with ordinary least squares regression, allows it to handle non-normal data effectively. This characteristic is one of the reasons for PLS-SEM's growing popularity in various fields, including business, marketing, and information systems (Hair et al., 2017).

Missing Data:

PLS-SEM can handle missing data more flexibly than CB-SEM. While it still prefers data to be missing at random (MAR) or entirely at random (MCAR), PLS-SEM can use pairwise deletion or mean substitution methods to manage missing data, making it more adaptable to datasets with incomplete information (Hair et al., 2017).

Aspect	CB-SEM	PLS-SEM
Sample Size Requirements	Typically, significant (200+ cases)	Smaller samples acceptable (100-200 cases)
Normality Assumptions	Assumes multivariate normality	No assumption of normality
Missing Data Handling	Requires MAR or MCAR; uses imputation or FIML	Flexible, uses pairwise deletion or mean substitution

Table 3 *Summary of Data Requirements Differences*

Therefore, CB-SEM and PLS-SEM have different data requirements that reflect their distinct estimation methods and objectives. CB-SEM requires larger sample sizes and normally distributed data to produce reliable estimates, while PLS-SEM is more flexible, accommodating smaller samples and non-normal data.

4. Handling of Measurement Error

Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM) use various approaches to address measurement error, which align with their specific goals and estimate techniques. Comprehending these distinctions is crucial for choosing the suitable approach according to a research investigation's distinct attributes and objectives. Measurement error is the discrepancy between a variable's actual and recorded values. This discrepancy may occur due to instrument defects, mistakes made by respondents, or external factors. CB-SEM and PLS-SEM use methodologies to mitigate measurement error but use different approaches.

CB-SEM Handling of Measurement Error

Explicit Modeling of Measurement Error:

CB-SEM explicitly models measurement error by incorporating it into the structural equation modeling framework. This approach allows researchers to separate the actual score variance from the error variance, leading to more accurate estimates of the relationships between latent variables (Kline, 2015).

Latent Variables and Indicators:

In CB-SEM, latent variables are conceptualized as underlying constructs measured by multiple observed indicators. Each indicator has an associated measurement error term explicitly included in the model. The measurement model is typically specified as follows:

$$X = \Lambda_x \xi + \delta$$
$$Y = \Lambda_y \eta + \epsilon$$

Where X and Y are vectors of observed variables, Λ_x and Λ_y are factor-loading matrices, ξ, and η are latent variables, and δ and ϵ are measurement error terms (Bollen, 1989).

Error Covariances:

CB-SEM allows for the specification of error covariances, which can capture the relationships between the measurement errors of different indicators. This flexibility enables more precise data modeling and the identification of potential sources of measurement error (Byrne, 2016).

Model Fit and Diagnostic Tools:

The goodness-of-fit indices used in CB-SEM, such as RMSEA, CFI, and TLI, help assess how well the model fits the observed data, including its measurement error components. Additionally, modification indices can be used to identify and address potential issues related to measurement error (Hu & Bentler, 1999).

PLS-SEM Handling of Measurement Error

Implicit Handling of Measurement Error:

PLS-SEM implicitly handles measurement errors through its estimation process. Instead of explicitly modeling measurement error, PLS-SEM focuses on maximizing the explained variance of the endogenous constructs, indirectly accounting for measurement error (Hair et al., 2017).

Formative and Reflective Measurement Models:

PLS-SEM can accommodate both formative and reflective measurement models. In reflective models, indicators are assumed to reflect the underlying latent variable, similar to CB-SEM. However, in formative models, indicators are considered to cause the latent variable and measurement error is implicitly accounted for by the model's structure (Chin, 1998).

Path Weights and Iterative Estimation:

PLS-SEM uses an iterative algorithm to estimate path weights and factor loadings, which helps mitigate the impact of measurement error. The algorithm adjusts the weights to maximize the explained variance, thereby reducing the influence of measurement error on the estimates (Henseler et al., 2015).

Composite Reliability and Average Variance Extracted:

PLS-SEM assesses the reliability and validity of the constructs using composite reliability (CR) and average variance extracted (AVE). These measures provide information about the proportion of variance in the indicators attributable to the latent construct versus measurement error (Hair et al., 2017).

Aspect	CB-SEM	PLS-SEM
Modeling Approach	Explicitly models measurement error	Implicitly accounts for measurement error
Measurement Model Specification	Latent variables measured by multiple indicators with explicit error terms	Can handle formative and reflective models with implicit error handling
Error Covariances	Allows specification of error covariances	It does not explicitly model error covariances
Diagnostic Tools	It uses goodness-of-fit indices and modification indices	Uses composite reliability and AVE for construct validation

Table 4 *Summary of Measurement Error Handling Differences*

Therefore, CB-SEM and PLS-SEM differ in their approaches to handling measurement errors. CB-SEM explicitly incorporates measurement error into the model, allowing for a detailed analysis of its impact on the observed data. Conversely, PLS-SEM implicitly addresses measurement error through its iterative estimation process, maximizing explained variance and ensuring construct reliability. These differences reflect each method's underlying philosophies and objectives, guiding researchers in choosing the appropriate approach based on their research needs and data characteristics.

5. Applicability

Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM) are applied in different research contexts based on their unique strengths and limitations. Understanding these applications helps researchers select the appropriate method for their research goals and contexts.

CB-SEM Applicability

Confirmatory Research:

CB-SEM is best suited for confirmatory research, where the primary goal is to test established theories. This method is used to verify whether the theoretical model fits the observed data. CB-SEM's rigorous statistical approach and emphasis on model fit make it ideal for studies that confirm hypotheses and validate theoretical constructs.

Standard Fields of Use:

CB-SEM is commonly used in the social sciences, psychology, and education, where theoretical models are well-developed and require robust testing. For example:

Social Sciences: Researchers use CB-SEM to test theories related to social behavior, cultural influences, and societal structures (Kline, 2015).

Psychology: In psychology, CB-SEM helps validate scales' psychometric properties, test cognitive and behavioral theories, and examine complex relationships among psychological constructs (Byrne, 2016).

Education: Educational researchers employ CB-SEM to validate educational measurement tools, test instructional theories, and evaluate the effectiveness of educational interventions (Schumaker & Lomax, 2016).

Characteristics of Research Suitable for CB-SEM:

- Well-defined theoretical framework
- Hypothesis testing
- Large sample sizes
- Normally distributed data
- PLS-SEM Applicability

- Exploratory Research:

PLS-SEM is ideal for exploratory research and situations where the primary goal is to predict and explain variance rather than to confirm a theoretical model. This method is advantageous when the research model is complex, the theoretical background is not well-established, or the primary interest is understanding key driver relationships and forecasting outcomes.

Standard Fields of Use:

PLS-SEM is frequently used in business, marketing, information systems, and management research. Its flexibility and ability to handle complex models with many indicators make it a popular choice in these fields. For example:

Business and Marketing: Researchers use PLS-SEM to explore customer satisfaction, brand loyalty, market segmentation, and consumer behavior (Hair et al., 2017).

Information Systems: In the field of information systems, PLS-SEM helps in understanding technology adoption, system usability, and the impact of information systems on organizational performance (Urbach & Ahlemann, 2010).

Management: Management researchers employ PLS-SEM to study leadership styles, organizational culture, innovation, and strategic decision-making processes (Henseler et al., 2015).

Characteristics of Research Suitable for PLS-SEM:

- Exploratory research objectives
- Prediction and explanation of variance
- Complex models with many constructs and indicators
- Small to medium sample sizes
- Non-normally distributed data

Aspect	CB-SEM	PLS-SEM
Research Focus	Confirmatory research, theory testing	Exploratory research, prediction, and variance explanation
Common Fields	Social sciences, psychology, education	Business, marketing, information systems, management
Theoretical Framework	Well-defined and established	Emerging or developing

Sample Size Requirements	Large	Small to medium
Data Distribution	Normally distributed	No assumption of normality

Table 5 Summary of Applicability Differences

CB-SEM and PLS-SEM serve different research purposes and are best suited for different stages of theoretical development. CB-SEM is ideal for confirmatory research that tests well-established theories. At the same time, PLS-SEM is better suited for exploratory research that seeks to predict and explain variance, especially in complex models.

Aspect	CB-SEM	PLS-SEM
Objective	Theory testing and confirmation	Prediction and theory development
Philosophy	Confirmatory	Exploratory
Model Fit Emphasis	The overall fit of the model to the data	Predictive power and explained variance
Key Fit Indices	Chi-square (χ^2), RMSEA, CFI, TLI, SRMR, GFI, AGFI	R-squared (R^2), Q-squared (Q^2), AVE, CR, SRMR
Approach to Fit	Confirmatory, goodness-of-fit indices	Exploratory, predictive relevance measures
Estimation Method	Maximum Likelihood Estimation (MLE)	Iterative algorithm combining PCA and OLS regression
Assumptions	Normality, large sample size	No assumption of normality, suitable for smaller samples
Handling of Measurement Error	Explicitly models measurement error	Implicitly accounts for measurement error
Measurement Model Specification	Latent variables measured by multiple indicators with explicit error terms	Can handle formative and reflective models with implicit error handling
Error Covariances	Allows specification of error covariances	It does not explicitly model error covariances

Diagnostic Tools	It uses goodness-of-fit indices and modification indices	Uses composite reliability and AVE for construct validation
Sample Size Requirements	Typically, significant (200+ cases)	Smaller samples acceptable (100-200 cases)
Normality Assumptions	Assumes multivariate normality	No assumption of normality
Missing Data Handling	Requires MAR or MCAR; uses imputation or FIML	Flexible, uses pairwise deletion or mean substitution
Research Focus	Confirmatory research, theory testing	Exploratory research, prediction, and variance explanation
Common Fields	Social sciences, psychology, education	Business, marketing, information systems, management
Theoretical Framework	Well-defined and established	Emerging or developing
Data Distribution	Normally distributed	No assumption of normality

Table 5 *Summary Table of Differences Between CB-SEM and PLS-SEM*

Conclusion

This study has provided a comprehensive comparison between Covariance-Based Structural Equation Modeling (CB-SEM) and Partial Least Squares Structural Equation Modeling (PLS-SEM), highlighting their distinct objectives, philosophies, data requirements, estimation methods, handling of measurement error, model fit indices, and applicability. By understanding these differences, researchers can decide which method to use based on their research objectives and data characteristics.

CB-SEM is best suited for confirmatory research, where the primary goal is to test established theories. It is commonly used in social sciences, psychology, and education, where well-developed theoretical models require rigorous testing. CB-SEM's emphasis on model fit and explicit handling of measurement error ensures that the hypothesized relationships among variables are supported by empirical evidence. In contrast, PLS-SEM is ideal for exploratory research and situations where the primary goal is to predict and explain variance. It is frequently used in business, marketing, information systems, and management research, where theoretical models may be less established, and the focus is on understanding key driver relationships and forecasting outcomes. PLS-SEM's flexibility in handling complex models with smaller samples and non-normal data makes it a powerful tool for exploratory studies.

Practical Implications for Research

Guidance on Method Selection: This study is a practical guide for researchers to choose between CB-SEM and PLS-SEM based on their research goals. Understanding each method's strengths and limitations helps select the most appropriate approach for theory testing or development.

Enhanced Methodological Rigor: This study enhances the methodological rigor of research designs by clarifying the data requirements and methodological nuances of CB-SEM and PLS-SEM. Researchers can ensure that their chosen method aligns with the characteristics of their data, leading to more reliable and valid results.

Improved Research Outcomes: Selecting the appropriate SEM technique can produce more accurate and meaningful research outcomes. For confirmatory research, using CB-SEM can validate theoretical constructs with high precision. For exploratory research, employing PLS-SEM can uncover new insights and relationships within the data.

Broader Applicability: The detailed comparison of CB-SEM and PLS-SEM makes this study relevant across various disciplines. Social sciences, business, psychology, education, and information systems researchers can benefit from the insights by applying the appropriate method to their research contexts.

In conclusion, this study provides valuable insights into the conceptual differences and practical applications of CB-SEM and PLS-SEM. Guiding researchers in selecting the appropriate study method contributes to advancing methodological rigor and generating reliable, valid, and impactful research findings.

References

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. doi:https://doi.org/10.1037/0033-2909.107.2.238

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* John Wiley & Sons.

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In J. Long, & K. Bollen, *Testing structural equation models* (pp. 136–162). Sage. doi:https://doi.org/10.1177/0049124192021002005

Byrne, B. M. (2016). *Structural Equation Modeling With AMOS 3rd Edition: Basic Concepts, Applications, and Programming, Third Edition.* New York: Routledge.

Cepeda, G., Roldán, J. L., Sabol, M., Hair, J., & Yee Loong Chong, A. (2024). Emerging opportunities for information systems researchers to expand their PLS-SEM analytical toolbox. *Industrial Management & Data Systems, 24*(6), 2230–2250. doi:https://doi.org/10.1108/IMDS-08-2023-0580

Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. In G. A. Marcoulides, *Modern methods for business research* (pp. 295–336). Lawrence Erlbaum Associates.

Hair Jr, J. F. (2020). Next-generation prediction metrics for composite-based PLS-SEM. *Industrial Management & Data Systems, 121*(1), 5–11. doi:https://doi.org/10.1108/IMDS-08-2020-0505

Hair, J. F., G. Tomas, H. M., Ringle , C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R.* Springer.

Hair, J. F., Hult, G. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM) (2nd ed.).* Sage Publications.

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. doi:https://doi.org/10.1108/EBR-11-2018-0203

Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review, 26*(2), 106–121.

Henseler, J., & Schuberth, F. (2024). Should PLS become factor-based, or should CB-SEM become composite-based? Both! *European Journal of Information Systems,* 1–13. doi:https://doi.org/10.1080/0960085X.2024.2357123

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. In A. Goldberger, & O. Duncan, *Structural equation models in the social sciences* (pp. 85–112). Seminar Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd ed.).* New York: Guilford Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling (4th ed.).* Guilford Publications.

Little, T. D. (2023). *Longitudinal Structural Equation Modeling.* New York: The Guilford Press.

Sabol, M., Hair, J., Cepeda, G., Roldán, J. L., & Yee Loong Chong, A. (2023). PLS-SEM in information systems: seizing the opportunity and marching ahead to adopt methodological updates. *Industrial Management & Data Systems, 123*(12), 2997–3017. doi:https://doi.org/10.1108/IMDS-07-2023-0429

Sarstedt, M., Adler, S. J., Ringle, C. M., Cho, G., Diamantopoulos, A., Hwang, H., & Liengaard, B. D. (2024). Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of Product Innovation Management, 42*(3), 1–17. doi:https://doi.org/10.1111/jpim.12738

Schumaker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling (3rd. Edition).* New Jersey: Lawrence Erlbaum Associates.

Sharma, P. N., Liengaard, B. D., Hair, J. F., Sarstedt, M., & Ringle , C. M. (2023). Predictive model assessment and selection in composite-based modeling using PLS-SEM: extensions and guidelines for CVPAT. *European Journal of Marketing, 57*(6), 1662–1677. doi:https://doi.org/10.1108/EJM-08-2020-0636

Trail, G. T., Kimb, Y. K., & Alfaro-Barrantes, P. (2024). A Critical Assessment for Sport Management Research: Comparing PLS-SEM and CB-SEM Techniques for Moderation Analysis Using Formative Measures. *Journal of Global Sport Management, 9*(1), 248–268. doi:https://doi.org/10.1080/24704067.2022.2098802

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10.

Urbach, N., & Ahlemann, F. (2010). Structural equation modeling in information systems research using Partial Least Squares. *Journal of Information Technology Theory and Application, 11*(2), 5–40.