Indonesian Journal of Business and Entrepreneurship, Vol. 11 No. 1, January 2025
Permalink/DOI: http://dx.doi.org/10.17358/IJBE.11.1.93

Available online at
http://journal.ipb.ac.id/index.php/ijbe

EVALUATING MACHINE LEARNING APPROACHES IN STRUCTURAL EQUATION MODELLING TO IMPROVE PREDICTIVE ACCURACY IN MARKETING RESEARCH

Chacha Magasi

Marketing Department, College of Business Education, Mwanza, Tanzania
P.O.Box 1968, Dar es Salaam, Dar es Salaam, Tanzania

Article history:

Received
13 August 2024

Revised
14 October 2024

Accepted
21 October 2024

Available online
24 January 2025

ABSTRACT

Background: This study aimed to fill a critical research gap by comparing traditional Structural Equation Modelling (SEM) with hybrid Bayesian-Machine Learning (ML) models in marketing research, focusing on the limited exploration of these advanced techniques.
Purpose: This study aimed to evaluate the effectiveness of integrating Bayesian SEM with advanced machine learning techniques to enhance predictive model performance, manage complex data structures, and improve marketing applications.
Design/methodology/approach: The study employed a systematic comparative research design to assess the predictive accuracy and robustness of traditional SEM in comparison to hybrid Bayesian-(Bayesian-ML) models. A rigorous review of 262 scholarly articles from major databases was conducted, with 23 studies meeting inclusion criteria to inform the model development and evaluation.
Findings/Result: The findings show that traditional SEM excels in theoretical modelling and interpretability but lacks predictive accuracy and robustness, which Bayesian SEM improves by using prior distributions. ML techniques further enhance predictive accuracy and robustness, while hybrid models combining Bayesian SEM with ML achieve the highest levels of both.
Conclusion: Adopting hybrid models can substantially enhance the predictive accuracy of marketing outcomes and the robustness of model analyses.
Originality/value (State of the art): This study contributes to knowledge by advancing methodological approaches through challenging existing data analysis paradigms, methods and approaches and therebefore offering practical guidance for future studies.

Keywords: accuracy, bayesian methods, hybrid models, machine learning, predictive, robustness, structural equation modelling (SEM)

How to Cite:
Magasi C. (2025). Evaluating Machine Learning Approaches in Structural Equation Modelling to Improve Predictive Accuracy in Marketing Research. Indonesian Journal of Business and Entrepreneurship (IJBE), 11(1), 93. https://doi.org/10.17358/ijbe.11.1.93

[1] Corresponding author:
Email: magasitza@gmail.com

INTRODUCTION

This study assesses the limitations of traditional Structural Equation Modelling (SEM) in managing complicated and non-linear complex data structures, which can lead to overfitting and reduced accuracy. The study evaluates how integrating the Bayesian SEM with machine learning techniques can enhance predictive accuracy and robustness. Studies covered are those focusing marketing research related to consumer behaviour, preferences, customer segmentation, and forecasting. The study starts by defining SEM as the sophisticated statistical framework that analyses complex and complicated relationships between observed and latent variables using both factor and path analysis. The major essence is to evaluate and precision test of causal relationships between the constructs (Harris & Gleason, 2022). The ability of SEM to handle multiple dependent relationships and latent variables makes it valuable in marketing research for understanding consumer behaviour, customer satisfaction and the market dynamics (Zyphur et al. 2023; Abbasi et al. 2024; Alzaydi, 2024). In addition, SEM facilitates the estimation of both direct and indirect effects when assessing the model fit through Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Tucker-Lewis Index (TLI) indices, consequently validating the theoretical models (Shi et al. 2019).

The traditional SEM consistently maintains precise modelling of the complex theoretical concepts and models particularly in marketing research because it has ability to systematically accommodate the latent variables (Ximénez et al. 2022).

Despite of its suitability in data analysis, the traditional SEM methods and approaches consists of several limitations. One of its limitations is heavily relies on Maximum Likelihood Estimation (MLE) which assumes normality. As a result, it may fumble with non-normal data distributions in studies which have large and complex data sets (Schweizer et al. 2023; Xu, 2024). Secondly, its predictive accuracy can be reduced and lead to overfitting whenever the model captures noise instead of the true patterns (Chattamvelli, 2024). Secondly, depending on traditional SEM robustness is challenging because it is extremely sensitive to sample size and model specification. Consequently, employing SEM to analyse the data from the small samples may result in unstable estimates and reduced generalizability (Chattamvelli, 2024; Ridler, 2023). Furthermore, the model may result in improper conclusions if the assumptions done are inaccurate because of the distorted omitted variables (Harring et al. 2017). Traditional SEM usually struggles with non-linear relationships and the complex interactions among the variables (Ridler, 2023). The need for a pre-established model specification and complex mathematical formulations may further limit SEM's practical utility and hence making it difficult to derive actionable recommendations by practitioners (Ridler, 2023; Hair & Alamer, 2022). So, to cover the stated gaps, this research intends to assess the effectiveness of adding other advanced machine learning techniques into a Bayesian SEM framework. The study intends to systematically evaluate how fusion of Bayesian and Machine Learning techniques to form hybridized models may enhance the predictive power and accuracy in highly complex datasets as compared to traditional SEM.

Though Traditional SEM is considered friendly in use and valuable, in most cases it faces challenges in predictive performance and robustness, primarily with complex or non-linear data structures (Lavelle-Hill et al. 2023). Studies report that when hybrid models use Bayesian methods and machine learning techniques to correctly capture the complex patterns, it offers superior stability and accuracy in data analysis (Bharadiya, 2023; Sansana et al. 2021; Brown, 2024; Patel & Gahletia, 2020). When using the datasets from a small sample size, Bayesian methods provide a solid alternative in enhancing model stability and accuracy and at the same time sufficiently addressing the non-linearity and overfitting (Bharadiya, 2023; Pazo et al. 2024). Integrating Bayesian methods with ML offers a transformative approach to tackle SEM shortcomings through incorporating the probabilistic reasoning and smartly pre-established distributions to effectively handling uncertainty and variability in the parameter estimates (Bharadiya, 2023). Bayesian SEM improves model robustness through posterior distributions, reducing overfitting and instability (Asosega et al. 2022). These said capabilities make Bayesian SEM especially effective for providing more accurate and stable parameter estimates in complex model datasets (Fan et al. 2016). Nevertheless, the application of Bayesian methods may computationally be intensive and needs careful specification of priors, which normally present practical challenges.

The ML techniques offer powerful algorithms capable of managing complex datasets with minimal prior specification (Bharadiya, 2023; Raschka, 2018). Techniques such as decision trees, neural networks, and random forests excel in capturing intricate non-linear relationships and interactions within large datasets (Taye, 2023). For example, random forests enhance predictive accuracy by aggregating the predictions of multiple decision trees, meanwhile neural networks, especially deep learning models, they build multiple layers of abstraction to effectively model complicated and complex relationships (Bharadiya, 2023; Taye, 2023). However, ML approaches may be prone to overfitting in high-dimensional settings, and their lack of interpretability can complicate the understanding of model dynamics and thus the need of hybrid models (Linardatos et al. 2020). Hybrid approaches fuse Bayesian SEM with machine learning techniques to skilfully and competently leverage the strengths of the said methodologies and at the same time addressing their specific limitations (Bharadiya, 2023; Kruschke, 2015). Bayesian methods adequately offer informative priors that smartly stabilize and sufficiently regularize machine learning models to accurately improve prediction precision and reliability. Concurrently, machine learning techniques refine and at the same time validate Bayesian models through employing advanced algorithms for model selection and performance evaluation. This results in an improved accuracy and robustness. Studies for example, Nosratabadi et al. (2020) reports that integrating Bayesian methods with ML significantly enhances predictive accuracy power, while Linardatos et al. (2020) highlights that hybrid models improve interpretability and robustness in complex data environments. Thus, hybrid approaches represent a significant advancement in modelling approaches and methodologies. Table 1 summarizes the key concepts, advantages, limitations, challenges and implications of each methodology.

Table 1. Summary of Literature Review on SEM, Bayesian Methods, ML Techniques, and Hybrid Approaches

Methodology	Key Concepts	Advantages	Limitations	Implications	Authors
Structural Equation Modelling (SEM)	Integration of factor and path analysis; evaluates model fit using RMSEA, CFI, TLI; handles latent variables.	Comprehensive framework for theoretical constructs; simultaneous estimation of multiple effects.	Sensitive to sample size; prone to overfitting; struggles with non-linear relationships.	Useful for validating theoretical models; may require large samples and additional methods to address non-linearity.	Grace & Bollen (2008); Fan et al. (2016); Shi et al. (2019); Nayyar (2022)
Bayesian Methods in SEM	Incorporates prior distributions and probabilistic reasoning; updates beliefs based on data.	Enhances robustness; handles small sample sizes well; provides uncertainty estimates.	Computationally intensive; requires careful prior specification; can be complex to implement.	Improves model stability and accuracy; useful in complex or small-sample scenarios.	Bharadiya (2023); Kruschke (2015); Fan et al. (2016)
Machine Learning Techniques	Includes decision trees, random forests, neural networks; excels in predictive tasks and capturing non-linear relationships.	Effective for large datasets; handles complex and non-linear relationships well; advanced predictive capabilities.	Prone to overfitting; often lacks interpretability. Computationally intensive; may require balancing complexities and interpretability.	Enhances predictive accuracy and model performance; useful for complex data but may require careful validation.	Raschka (2018); Taye (2023); Bharadiya (2023); Linardatos et al. (2020); Rudin et al. (2022)
Hybrid Approaches	Combines Bayesian SEM with ML techniques; integrates strengths of both methodologies.	Leverages Bayesian stability with ML predictive power; improves accuracy and robustness.	Computationally intensive; may require balancing complexities and interpretability.	Offers a robust framework for addressing limitations of individual methods; suitable for complex and dynamic data.	Bharadiya (2023); Kruschke (2015); Ma & Sun (2020); Linardatos et al. (2020); Richter & Tudoran (2024)

METHODS

This study employs a systematic comparative research design to evaluate the predictive accuracy and robustness of traditional SEM compared to hybrid Bayesian-ML models. The study examines how integrating Bayesian SEM with advanced ML techniques can enhance predictive accuracy and model robustness. The study utilized AND to integrate key concepts such as 'ML' and 'SEM,' OR to include related terms like 'predictive analytics,' and NOT to eliminate extraneous topics. A total of 262 peer reviewed articles were retrieved from three academic databases: 101 from Web of Science, 89 from Scopus, and 72 from ABI/INFORMS. After the elimination of duplicates, 209 articles remained for further evaluation. A subsequent title and abstract review led to the exclusion of 94 studies, resulting in 115 articles advancing to the next stage of scrutiny. These articles were assessed against predefined inclusion criteria, which stipulated that they be empirical in nature, contain relevant data, provide full-text access, be publicly accessible, and specifically address marketing issues. Ultimately, 23 articles fulfilled the inclusion criteria and were selected for inclusion in the final analysis. The process map is indicated in Figure 1.

The evaluation of the models employed well-defined metrics, including predictive accuracy assessed through the sample size Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Standardized Root Mean Square Residual (SRMR), and R-squared values. SEMs are usually evaluated through a range of model fit indices and relevant metrics (Asosega, et al. 2022). RMSEA evaluates the degree of model misfit such that 0 shows a perfect fit while relatively higher values indicating poor fit. It is normally useful for detecting erroneous models and is less affected by sample size when compared to chi-square test, with an acceptable value being less than 0.06 (Asosega, et al. 2022). In a similar vein, SRMR measures the variation between observed and predicted correlations. If its value less than 0.09 means a good model fit. The CFI assesses how well the model explains the variance in the covariance structure data. Its values range from 0 to 1, where a value of 0.95 or higher is accepted. CFI is normally less sensitive to sample size than chi-square test (Asosega, Iddrisu, Tawiah, Opoku, & Okyere, 2022). The TLI can improve the limitations of Normed Fit Index and is not influenced by sample size. Its score of 0.90 or above is acceptable for SEM models. The stability, scalability, and interpretability of models were evaluated based on their relative performance across marketing contexts, clarity, conditions and ability to manage complex data..

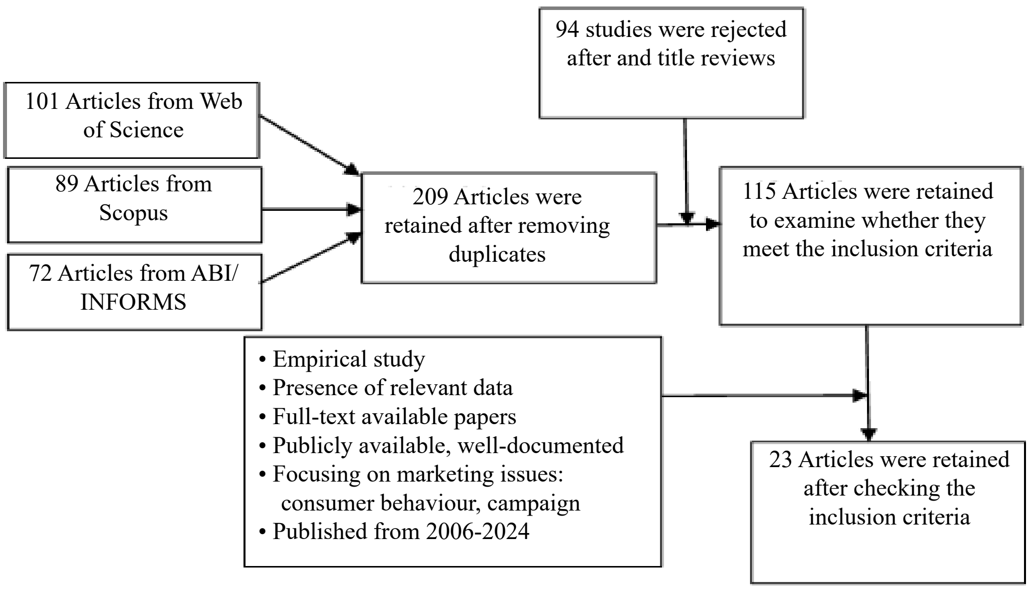

Figure 1. Research design

RESULTS

Findings on Predictive Modelling Techniques

This part presents the findings on various predictive modelling approaches as outlined in Table 2. Although traditional SEM is the foundational and clear, it has limited predictive accuracy and precision, with higher RMSE and MAE. Also, it is sensitive to sample size and overfitting. The SEM improves Chi-Square values with larger sample sizes, although; RMSEA show less sensitivity to variations in sample size. The robustness of SEM is distinguished as moderate, predominantly influenced by the underlying assumptions and the sample size (Pazo et al. 2024; Gerassis et al. 2024). Besides, SEM faces scalability limitations for larger sample size, which raises the risk of overfitting. Notwithstanding these limitations, SEM remains highly interpretable due to its well-established theoretical framework. Empirical metrics consider RMSEA of 0.06 as acceptable fit, while the threshold of acceptable CFI is considered 0.95. TLI stands at 0.90, considered acceptable, and RMSE is at 0.12, indicating a moderate error level (Asosega et al. 2022).

Table 2. Findings on predictive modelling techniques: insights from SEM to hybrid approaches

Model Type	Sample Size	Model Robustness	Scalability	Interpretability	Metrics Used	Authors
Structural Equation Modelling (SEM)	Larger sample sizes generally enhance Chi-Square values, though RMSEA remains less sensitive.	Moderate robustness; highly contingent on assumptions and sample size.	Limited scalability for larger samples, with a risk of overfitting.	Highly interpretable due to established principles.	RMSEA = 0.06 (acceptable fit), CFI = 90 or 0.95 (approaching threshold), TLI = 0.90 (acceptable), RMSE = 0.12 (moderate error).	Asosega et al. (2022); Pazo et al. (2024); Gerassis et al. (2024)
Bayesian Model	The significance of Chi-Square diminishes as Bayesian models utilize prior distributions.	Robust management of parameter uncertainty, moderately influenced by sample size.	Moderate scalability, with improved performance as sample size increases compared to SEM.	Interpretability depends on chosen prior distributions and model structure.	RMSEA cutoff point: upper limit = 0.08, lower limit = 0.05 (RMSEA϶[0.05,0.08]), CFI ≥ 0.90 or 95, TLI > 0.90, RMSE = 0.09 (lower error than SEM).	Hoofs et al. (2017); (2013); Khalid (2024); Doss-Gollin & Keller (2023)
ML	Model accuracy significantly improves with larger datasets, enhancing precision.	Highly robust and adaptive, especially with complex datasets.	Exceptionally scalable, adept at handling vast datasets.	Interpretability can be complex, depending on the algorithm used.	RMSE = 0.07 (lower error rate), Accuracy = 85%, F1-Score = 0.88, Chi-Square becomes less relevant.	Hoofs et al. (2017); Kruschke (2015); Rudin et al. (2022)
Bayesian-ML (Hybrid model)	Larger sample sizes enhance both prediction accuracy and robustness, reducing the relevance of Chi-Square.	Highly robust, integrates Bayesian regularization for model adaptability.	Highly scalable, effective for large, intricate datasets.	Varies with the ML techniques applied.	RMSEA = 0.035 (excellent fit), CFI = 0.96 (exceeds threshold), TLI = 0.93, RMSE = 0.05 (lowest error rate).	Asosega et al. (2022); Sansana et al. (2021); Bharadiya (2023); Hoofs et al. (2017)

Bayesian SEM offers improved accuracy, precision and robustness with lower RMSE and MAE, despite facing challenges with non-linear data and complexity. Consequently, the significance of Chi-Square diminishes as these models use the prior distributions, and thus effectively managing parameter uncertainty. These models display moderate robustness, with some influence from the sample size (Khalid, 2024; Doss-Gollin & Keller, 2023). The scalability of Bayesian Models is similarly moderate, with performance improvement as the sample size increases. Interpretability in Bayesian frameworks is subject to the chosen prior distributions and model structure. Empirical findings indicate that the established RMS cutoff points are at 0.08 with upper limits at 0.05 with lower limits and hence delineating a range of acceptable fit. TLI values more than 0.90 are deemed acceptable while those which are above 0.95 signify excellent fit. Besides, CFI values greater than 0.90 indicate acceptable performance while that with values above 0.95 categorized as excellent (Guenther et al. 2023; Linardatos et al. 2020). Generally, RMSEA values which are less than 0.08 are considered acceptable while those less than 0.05 show excellent fit (Gelman et al. (2013).

The ML methods achieve superior predictive accuracy and handle complex and complicated data effectively though they lack interpretability. The analysis of ML methodologies shows that model accuracy tends to significantly improves with larger data structures, consequently enhancing predictive precision. The machine learning models show relatively high robustness and smart adaptability when handling complex and complicated datasets (Kruschke, 2015; Rudin et al. 2022). Besides, machine learning techniques exhibit exceptional scalability, and they successfully manage the large volumes of data. Nonetheless, interpretability may be complex and varies depending on the specific algorithms employed. Empirical findings indicate that RMSE of 0.07 demonstrates a lower error rate with an accuracy rate of 85% and an F1-Score of 0.88. Thus, Chi-Square metrics become less relevant (Murphy, 012; Neeraj, 2020).

Hybrid approaches which incorporate both Bayesian and ML techniques, usually deliver the highest accuracy and robustness while improving interpretability and hence addressing the limitations of each individual method. The findings related to Bayesian-ML indicate that larger sample sizes enhance both prediction accuracy and robustness. As result, they tend to diminish the significance of Chi-Square metrics. This model is further characterized by high robustness, as it integrates Bayesian regularization for adaptability (Sansana et al. 2021; Bharadiya, 2023). Additionally, Bayesian-ML demonstrates excellent scalability and thus proving effective for large and intricate data structures. Interpretability usually varies based on the specific applied ML techniques. Empirical metrics for this hybrid model indicate that the RMSEA of 0.035 is considered an excellent fit; the CFI of 0.96, surpassing the acceptable threshold; a TLI of 0.93; and the RMSE of 0.05, reflecting the lowest error rate among the assessed models (Asosega et al. 2022).

Predictive Accuracy

Traditional SEM is foundational in theoretical research because of its capacity to estimate direct and indirect effects and assess model fit using indices such as RMSEA, CFI and TLI. Nonetheless, these models usually exhibit limitations in predictive accuracy. This is evidenced by higher MAE and RMSE values as compared to Bayesian SEM and ML approaches. This is because SEM relies on linear relationships, sensitive to sample size, and lacks the ability to handle complex and complicated non-linear data as effectively as more advanced predictive models including Bayesian SEM, random forests and neural networks do in offering very superior predictive accuracy and robustness. This signifies that while Traditional SEM is preferably more effective for theoretical research and model evaluation, its limitation in predictive accuracy is likely to affect its practical utility. The findings further indicate that while traditional SEM may fit well the theoretical constructs. Furthermore, its performance in forecasting and personalized recommendations becomes suboptimal because of its reliance on linear relationships and sensitivity to sample size. Thus, because of its reliance on linear relationships and sensitivity to sample size, it makes it less suitable for forecasting and personalized recommendations when compared to Bayesian SEM and ML methods. These findings are consistent with research by Pazo, Gerassis, Araújo, Antunes, and Rigueira (2024) that traditional SEM methods may not fully capture complex data patterns and therefore impacting their predictive power. In the similar vein, Khalid (2024) highlights that because of using SEM model to analyse quantitative data gathered from a diverse sample population, the small sample size of 509 responses was revealed as reducing the statistical power and potentially leading to either inconclusive or inaccurate results. Moreover, in small-sample contexts like marketing research, it is important to carefully evaluate imputation techniques, including median replacement, expectation-maximization algorithm, and winsorization of outliers, to effectively manage the missing data or outliers (Guenther, Guenther, Ringle, & Zaefarian, 2023). The SEM models normally rely on model fit indices such as the Chi-Square test, p-values, and standard errors, which to large extent are significantly influenced by the sample size, with smaller samples regularly leading to imprecise estimates and poor model fit (Asosega, Iddrisu, Tawiah, Opoku, & Okyere, 2022).

Since Bayesian SEM approaches uses prior distributions to regularize estimates and manage uncertainty, they show much more improved predictive accuracy when compared to traditional SEM. Bayesian methods usually achieve lower MAE and RMSE values which are less sensitive to sample size, an indication in enhanced performance in forecasting and personalized recommendations. The averaging Bayesian model predicts future outcomes by considering all models weighted by their posterior probabilities rather than relying heavily on the best-fitting model or sample size (Kruschke, 2015). This approach inherently accounts for model complexity by compensating for the tendency of complex and complicated models to fit noise. The approach achieves the said outcomes by spreading prior probabilities thinly across the larger parameter spaces, and therefore balancing model fit and complexity to consistently avoiding overfitting the random data variations. For example, in coin bias scenarios the Bayesian method ensures that biases are primarily factored into the prediction to sufficiently enhancing the accuracy (Kruschke, 2015). Thus, the Bayesian SEM's ability to systematically fuse prior distributions to regularize estimates and accurately manage prevailing uncertainty, results in enhancing its predictive accuracy and thus making it much superior to traditional SEM for forecasting and personalized recommendations. The results are in consistent to Doss-Gollin and Keller (2023) and McNeish (2016) observations who highlight that Bayesian SEM manages the complex models and small sample sizes much more effectively and accurately than traditional SEM. Similarly, the Bharadiya (2023) revealed that Bayesian ML, as a unique subfield of ML, utilizes Bayesian principles and probabilistic models to improve the learning process. However, while Bayesian SEM indicates better predictive performance, it might not entirely exploit the advanced pattern recognition capabilities of machine learning. Thus, although Bayesian SEM demonstrates superior predictive performance, it may not leverage advanced pattern recognition capabilities of ML techniques. Also, Pazo et al (2024) highlights that Bayesian methods alone can hardly handling highly non-linear data.

The machine learning techniques such as decision trees, random forests, support vector machines and neural networks, consistently exceed both traditional SEM and Bayesian SEM in their predictive accuracy. The aforementioned methods are capable of identifying complex, non-linear relationships and handling large datasets, resulting to the lowest MAE and RMSE values. Therefore, machine learning techniques such as decision trees, random forests, support vector machines and neural networks outperform both traditional and Bayesian SEM in predictive accuracy. They effectively recognize complex and non-linear relationships and thereafter competently manage large datasets which results in scoring the lowest MAE and RMSE values. The findings from this study spotlight that machine learning models such as LIME and SHAP significantly primarily enhance marketing disciplines and models by enhancing predictive accuracy, interpretability, model robustness and ultimately strategic implications. For instance, customer segmentation and personalized marketing leads in up to a 20 percent increase in precision, resulting in enhanced marketing outcomes such as a 15 percent reduction in customer acquisition costs (Linardatos, Papastefanopoulos, & Kotsiantis, 2020). Furthermore, Neeraj (2020) and Rudin, et al. (2022) revealed that machine learning is superior in managing intricate data patterns. Nonetheless, while machine learning approaches give significant improvements in prediction, they may regularly fall short of theoretical grounding and precision interpretability (Linardatos, Papastefanopoulos & Kotsiantis,2020). Thus, despite their significant improvements in the predictive performance, ML approaches might be deficient in the theoretical grounding and interpretability that SEM offers.

As discussed earlier the Hybrid models are obtained after integrating Bayesian methods with machine learning techniques. The hybridized models significantly achieve the highest predictive accuracy among all assessed approaches in this study. The models achieve substantial lower MAE and RMSE values and therefore emphasizing their superior performance particularly in forecasting and personalized recommendations. Therefore, the hybrid models are significantly graded as having the highest predictive accuracy and at the same time substantially lowering MAE and RMSE values. This finding is consistent to Bharadiya (2023) and Sansana, et al. (2021), who that hybrid approaches leverage the strengths of both Bayesian methods and ML methodologies by achieving the highest predictive accuracy and significantly reducing MAE and RMSE values. Theoretically, blending Bayesian methods with ML techniques can address the limitations of each and improve the predictive power and accuracy. Practically, the approach enhances the performance in forecasting and personalized recommendations, suggesting that future studies should focus on optimizing the aforementioned models for diverse applications.

Model Robustness

Traditional SEM models unveil variable robustness based on the sample size and model complexity. Sensitivity analyses indicate that these models may be unstable with small sample sizes, resulting to unreliable parameter estimates and the reduced generalizability. Besides, McNeish (2016) reports that SEM is not only susceptible to overfitting but also likely missing to perform consistently across various sample size and datasets and hence challenges their efficacy in the dynamic marketing environment. Sensitivity to small sample sizes and susceptibility to overfitting weaken the generalizability and effectiveness of traditional SEM models, especially in these dynamic contexts and environment.

Bayesian SEM approach enhances robustness when compared to traditional SEM approaches through using prior distributions to regularize estimates and manage uncertainty. This method reduces the risk of overfitting and significantly improves parameter stability, especially in small sample scenarios or complicated complex models. Therefore, Bayesian SEM significantly improves robustness when compared to traditional SEM by leveraging prior distributions to stabilize estimates and manage uncertainty. Consequently, it reduces overfitting and substantially enhancing performance in small sample or complex scenarios. Bharadiya (2023) and Doss-Gollin and Keller (2023) emphasize that Bayesian SEM is flexible and able to adapt to varying data conditions. However, in spite of aforementioned advantages, Bayesian SEM faces a number of challenges related to computational complexity and therefore the need for careful specification of prior distributions.

The ML techniques such as random forests and neural networks, provide superior robustness when compared to traditional SEM and Bayesian SEM. Random forests tendto enhance stability through an ensemble of decision trees, and substantially reducing susceptibility to overfitting. Neural networks offer robustness by modelling complex interactions in big datasets. This signifies that ML techniques generally offer greater robustness than traditional SEM and Bayesian SEM through enhancing stability, reducing overfitting, and effectively modelling the complex interactions in large datasets. With regard to model robustness, ML techniques mitigate bias and have been instrumental in promoting fairness, especially in pricing strategies and advertisement delivery. Consequently, it has led to 10% reduction in discriminatory practices against the marginalized consumer groups (Linardatos, Papastefanopoulos, & Kotsiantis, 2020). The improvement assures that marketing strategies are both equitable and effective and significantly reinforce trust and compliance within diverse market contexts. Nonetheless, the computational demands and potential overfitting in neural networks are concerns highlighted on ML techniques (Neeraj, 2020; Raschka,2018). In spite of their robustness, ML approaches face several challenges including high computational demands and overfitting, and therefore requiring either alternative solutions or enhancements.

Hybrid models that integrate Bayesian SEM with ML techniques shows the highest level of robustness than ever. The Bayesian methods contribute to model regularization and thus reducing overfitting. Moreover, ML algorithms substantially enhance adaptability to complex datasets. The combination ensures consistent and accuracy performance across diverse data conditions. As a result, it addresses the robustness challenges that are being faced by traditional SEM and Bayesian SEM. The integration of aforementioned methodologies provides a comprehensive solution to the issues of model stability and adaptability (Bharadiya, 2023 & Sansana, et al. 2021). Theoretically, integrating Bayesian SEM with ML techniques significantly enhances robustness by merging Bayesian regularization with machine learning's adaptability. Practically, this hybrid approach offers a robust and relatively stable solution for diverse data conditions and hence optimizing performance across various applications.

Interpretability

The findings indicate that SEM models are normally appreciated due to their interpretability by consistently offering clear paths and relationships between the observed variables and the latent variables. This clarity is valuable since it helps in the derivation of actionable and decision. Scholars such as Linardatos, Papastefanopoulos, and Kotsiantis (2020) and Rudin, et al. (2022) assert that the directly representation of theoretical constructs in SEM models makes them more highly accessible for interpretation. Nonetheless, as the complexity of SEM models increases with multiple latent variables, interpretability becomes more challenging.

Bayesian SEM significantly enhance predictive accuracy by employing prior distributions to regularize estimates and manage uncertainty. This results to more reliable and precise predictions. Nevertheless, this method unveils challenges in interpretability because of using complex prior distributions and posterior estimation can obscure the clarity of relationships between variables when compared to traditional SEM. Linardatos, Papastefanopoulos, and Kotsiantis (2020) and Rudin, et al. (2022) argue that Bayesian SEM's increased complexity can obscure the clarity of relationships between variables. If clarity is obscured, it becomes more difficult to generate actionable insights and make informed decisions by relying on the model's outputs.

The ML models are highly effective in managing complex data and giving high predictive accuracy due to their ability to discern complex patterns and relationships within large datasets. Furthermore, interpretability has emerged as an important aspect, with methods such as SHAP being cited in over 12,000 research papers for their ability to offer granular insights into consumer behaviour. These insights allow researchers to better understand the drivers of purchasing decisions, optimize product recommendations, and refine loyalty programs (Linardatos, Papastefanopoulos, & Kotsiantis, 2020). Despite these advancements, only 10-15% of marketing strategies can fully incorporate interpretability methods, indicating the significant untapped potential for their broader application. However, the inherent algorithmic complexity of ML models often makes interpretability challenging, obscuring the understanding of how predictions are derived (Bharadiya, 2023 ; Lavelle-Hill, et al. 2023 ; Linardatos, et al. 2020). The decision-making processes in these models can be opaque, making it difficult to understand how predictions are generated, and which features are most influential. This means that the complexity of ML models often makes it difficult to understand how predictions are made, and which features are most influential, potentially reducing trust and usability. Linardatos, Papastefanopoulos, and Kotsiantis (2020) and Rudin, et al. (2022) emphasize that while techniques like feature importance analysis and partial dependence plots can provide some level of interpretability, they do not fully resolve the challenges of understanding complex ML models. This limitation highlights the trade-off between predictive accuracy and interpretability in ML approaches.

Hybrid models that integrate Bayesian SEM with ML techniques provide a comprehensive approach to interpretability by combining Bayesian regularization with advanced pattern recognition. The Bayesian component contributes to understanding model uncertainty and parameter regularization, while ML algorithms provide advanced predictive capabilities (Bharadiya, 2023 ; Doss-Gollin & Keller, 2023 ; Pazo, et al. 2024). Practically, this means that hybrid models are beneficial because they not only improve prediction accuracy through advanced machine learning techniques but also offer better interpretability and understanding of model uncertainty through Bayesian methods. This combination makes the models more useful for decision-making and insights in complex data scenarios. Hybrid models enhance interpretability by using techniques such as feature importance analysis and partial dependence plots. Feature importance analysis identifies the most influential variables in the model's predictions, while partial dependence plots demonstrate how variations in specific features affect the predictions, holding other factors constant. These methods collectively clarify how different variables contribute to the model's outcomes. Despite these efforts, balancing interpretability with the sophisticated capabilities of hybrid models remains a challenge, as supported by Linardatos, Papastefanopoulos, and Kotsiantis (2020) and Rudin, et al. (2022) by Ribeiro et al. (2022). The ongoing development of interpretability techniques is essential for making hybrid models more accessible and useful in practical applications.

Managerial Implications

From a theoretical standpoint, this research makes a significant contribution to the field of SEM by demonstrating how the integration of Bayesian methods with ML can address the traditional limitations associated with SEM, particularly concerning predictive accuracy and model stability. Practically, the study equips marketing practitioners with robust analytical tools that facilitate deeper data insights, enabling informed and strategic decision-making. The improved accuracy and robustness of the hybrid models enable marketers to effectively and accurately develop, formulate and implement marketing strategies that are more precisely and well aligned with consumer needs and preferences, and finally fostering improved customer engagement and superior marketing outcomes. Despite hybrid Bayesian-Machine learning models demonstrating significant advancements in analytical capabilities, they have limitations. The complexity of hybrid Bayesian-ML models brings in the substantial computational resources and notably prolonged processing times. As a result, they are potentially limiting their practical applicability for research with large datasets or real-time analytical environments. Also, despite being advanced in predictive accuracy, the interpretability of hybrid Bayesian-ML models remains a major challenge. Therefore, it is important for organizations to ensure access to high-quality and unbiased data. This is because the accurate performance of hybrid models depends on the accuracy, completeness and integrity of the input data. Avoiding potential biases in the data collection process critically enhance the reliability and the generalizability of hybrid Bayesian-ML models.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This study evaluates the efficacy of incorporating Bayesian SEM with advanced machine learning techniques to enhance predictive accuracy and model robustness in marketing research. The findings give the distinct strengths and limitations built-in to each methodology. Traditional SEM exhibits commendable proficiency in theoretical modelling and interpretability. Nevertheless, it sometimes demonstrates deficiencies in predictive accuracy because of relying on linear assumptions and very sensitive to sample size. Conversely, Bayesian SEM alleviates some of the stated limitations by utilizing prior distributions, and hence significantly enhancing the predictive accuracy. These models usually leverage and enhance the strengths of both Bayesian statistical methods and ML algorithms and hence offering improved accuracy in managing uncertainty and accurately making probabilistic predictions. The models are very valuable in applications needing precise interpretability and robustness including personalized recommendations and predictive analytics. Notwithstanding, Bayesian SEM struggles with the complexities of managing complex datasets while attempting to systematically maintain interpretability. Machine learning techniques are characterized by their ability in modelling non-linear relationships. Thus, they offer much more superior predictive capabilities despite of regularly lacking the interpretability necessary for deriving actionable insights. The findings indicate that hybrid models, which integrate Bayesian SEM with ML techniques, successfully leverage the strengths of both approaches and paradigms. Nevertheless, these models are practically computationally intensive and essentially require significant expertise for effective implementation. Moreover, incorporation of Bayesian methods and machine learning may occasionally result in complexity in model specification and evaluation and therefore causing complications in practical deployment. This study is original as it comprehensively comparatives analysis of SEM and hybrid Bayesian-ML models with unique findings that the latter approach significantly improves the predictive capabilities and at the same time challenging existing assumptions on the limitations of SEM and the potential of machine learning. In addition, this study critically assesses the routine reliance on traditional SEM in marketing studies and thus offering the novel perspectives and actionable insights for refining both theoretical frameworks and practical applications. Based on the findings, hybrid models should be adopted by marketers and researchers as a means of achieving more accurate and robust data analyses.

Recommendations

This study assessed the limitations of traditional Structural Equation Modelling (SEM) in managing complicated and non-linear complex data structures, which can lead to overfitting and reduced accuracy. The study further evaluated how integrating the Bayesian SEM with machine learning techniques can enhance predictive accuracy and robustness. Future study may investigate how further development of hybrid Bayesian-ML models development of methodologies can enhance their transparency and interpretability and hence rendering them more accessible and actionable. In addition, empirical investigations to explore the application of these models across diverse marketing contexts and in other industries to further substantiate their effectiveness. Additionally, the fusion of advanced methodologies, for example deep learning and ensemble methods with Bayesian and machine learning methods may further promise augmenting predictive accuracy and robustness. The purpose is to refine the analytical tools be available to marketers and significantly contribute to improvement of the ongoing marketing research methodologies.

REFERENCES

Abbasi, A. Z., Rather, R. A., Hooi Ting, D., Nisar, S., Hussain, K., Khwaja, M. G., & Shamim, A. (2024). Exploring tourism-generated social media communication, brand equity, satisfaction, and loyalty: A PLS-SEM-based multi-sequential approach. Journal of Vacation Marketing, 30(1), 93-109. https://doi.org/10.1177/13567667221118651.

Alzaydi, Z. (2024). Examining the mediating effect of multi-channel integration quality in the relationship with service quality, customer satisfaction and customer loyalty in the Saudi banking sector. Management & Sustainability: An Arab Review, 3(2), 132-149. https://doi.org/10.1108/MSAR-12-2022-0061.

Asosega, K. A., Iddrisu, W. A., Tawiah, K., Opoku, A. A., & Okyere, E. (2022). Comparing bayesian and maximum likelihood methods in structural equation modelling of university student satisfaction: an empirical analysis. Education Research International, 1-12. https://doi.org/10.1155/2022/3665669.

Bharadiya, J. P. (2023). A review of Bayesian machine learning principles, methods, and applications. International Journal of Innovative Science and Research Technology, 85, 2033-2038.

Brown, I. (2024). Mastering Marketing Data Science: A Comprehensive Guide for Today's Marketers. John Wiley & Sons.

Chattamvelli, R. (2024). Correlation in engineering and the applied sciences: Applications in R. Springer Nature. Springer. https://books.google.co.tz/books?hl.

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in finance (Vol. 1170). New York, NY, USA: Springer International Publishing. https://link.springer.com/book/10.1007/978-3-030-41068-1.

Doss-Gollin, J., & Keller, K. (2023). A subjective Bayesian framework for synthesizing deep uncertainties in climate risk management. Earth's Future, 11(1), 1-19. https://doi.org/10.1029/2022EF003044.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu. R., S., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. Ecological Processes, 5(19), 1-12. https://link.springer.com/article/10.1186/s13717-016-0063-3.

Grace, J. B., & Bollen, K. A. (2008). Representing general theoretical concepts in structural equation models: the role of composite variables. . Environmental and Ecological Statistics, 15, 191-213. https://doi.org/10.1007/s10651-007-0047-7.

Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. Research Methods in Applied Linguistics, 1(3), 1-16. https://doi.org/10.1016/j.rmal.2022.100027.

Hamaker, E. L., Asparouhov, T., & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. The handbook of structural equation modeling (31 ed.). Los Angeles: Guilford Press.

Harring, J. R., McNeish, D. M., & Hancock, G. R. (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. Psychological Methods, 22(4), 616-631. https://psycnet.apa.org/buy/2017-55458-002.

Harris, J. E., & Gleason, P. M. (2022). Application of path analysis and structural equation modeling in nutrition and dietetics. Journal of the Academy of Nutrition and Dietetics, 122(11), 2023-2035. https://doi.org/10.1016/j.jand.2022.07.007.

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. Advances in methods and practices in psychological science, 2(1), 55-76. https://doi.org/10.1177/2515245919826527.

Khalid, B. (2024). Evaluating customer perspectives on omnichannel shopping satisfaction in the fashion retail sector. Heliyon, 10(16), 1-17. .

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. Artificial Intelligence Review, 26, 159-190.

Kruschke, J. (2015). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2 ed.). Bloomington: Elsevier.

Lavelle-Hill, R. E., Smith, G., & Murayama, K. (2023). Machine Learning Meets Traditional Statistical Methods in Psychology: Challenges and Future Directions. OSF Preprints, 1-22. .

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. Entropy, 23(1), 1-11. https://doi.org/10.3390/e23010018.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. International Journal of Research in Marketing, 37(3), 481-504. https://doi.org/10.1016/j.ijresmar.2020.04.005.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. Structural Equation Modeling: A Multidisciplinary Journal, 23(5), 750-773. https://doi.org/10.1080/10705511.2016.1186549.

Nayyar, V. (2022). Reviewing the impact of digital migration on the consumer buying journey with robust measurement of PLS-SEM and R Studio. Systems Research and Behavioral Science, 39(3), 542-556. https://doi.org/10.1002/sres.2857.

Neeraj, K. N.-2. (2020). A review on machine learning (feature selection, classification and clustering) approaches of big data mining in different area of research. Journal of critical reviews, 7(19), 2610-2626.

Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., & Gandomi, A. H. (2020). Data science in economics: comprehensive review of advanced machine learning and deep learning methods. Mathematics, 8(10), 1-13. https://doi.org/10.3390/math8101799.

Patel, P., & Gahletia, S. (2020). Deriving Actionable Insights from Big Data to Enhance Customer Experiences Across the Consumer Journey. International Journal of Responsible Artificial Intelligence, 10(8), 1-9. https://neuralslate.com/index.php/Journal-of-Responsible-AI/article/view/61.

Pazo, M., Gerassis, S., Araújo, M., Antunes, I. M., & Rigueira, X. (2024). Enhancing water quality prediction for fluctuating missing data scenarios: A dynamic Bayesian network-based processing system to monitor cyanobacteria proliferation. Science of the of the Total Environment, 927, 1-13. https://doi.org/10.1016/j.scitotenv.2024.172340.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. . arXiv preprint arXiv:1811.12808., 1-49. https://doi.org/10.48550/arXiv.1811.12808.

Richter, N. F., & Tudoran, A. A. (2024). Elevating theoretical insight and predictive accuracy in business research: Combining PLS-SEM and selected machine learning algorithms. Journal of Business Research, 173, 1-18. https://doi.org/10.1016/j.jbusres.2023.114453.

Ridler, I. (2023). Applications of regularised structural equation modelling to psychometrics and behavioural genetics. Doctoral dissertation,. London: King's College London.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistic Surveys, 16, 1-85. https://doi.org/10.1214/21-SS133.

Sansana, J., Joswiak, M. N., Castillo, I., Wang, Z., Rendall, R., Chiang, L. H., & Reis, M. S. (2021). Recent trends on hybrid modeling for Industry 4.0. Computers & Chemical Engineering, 151, 1-14. https://doi.org/10.1016/j.compchemeng.2021.107365.

Schweizer, K., DiStefano, C., & French, B. (2023). A maximum likelihood approach for asymmetric non-normal data using a transformational measurement model. Frontiers in Applied Mathematics and Statistics, 9, 1-9. https://doi.org/10.3389/fams.2023.1095769.

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. Educational and psychological measurement. Educational and psychological measurement, 79(2), 310-334.doi.org/10.1177/0013164418783530.

Shmueli, G., Ray, S., Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. Journal of business Research, 69(10), 4552-4564. doi.org/10.1016/j.jbusres.2016.03.049.

Singh, V., Pencina, M., Einstein, A. J., Liang, J. X., Berman, D. S., & Slomka, P. (2021). Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. Scientific reports, 11(1), 1-8. https://doi.org/10.1038/s41598-021-93651-5.

Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers, 12(5), 1-16. https://doi.org/10.3390/computers12050091.

Xu, Y. (2024). Contributions to Parametric and Semiparametric Structural Equation Modeling Methods With Applications in the Biomedical Sciences (Doctoral dissertation). Georgetown University.

Zyphur, M. J., Bonner, C. V., & Tay, L. (2023). Structural equation modeling in organizational research: The state of our science and some proposals for its future. Annual Review of Organizational Psychology and Organizational Behavior, 10(1), 495-517. doi.org/10.1146/annurev-orgpsych-041621-031401.