DOI: http://dx.doi.org/10.26483/ijarcs.v16i1.7176

Volume 16, No. 1, January-February 2025

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Sequential Bayesian SEM for Task Technology Fit

Rabeah Samimi
Graduate School of Engineering and Science
University of the Ryukyus
1 Senbaru Nishihara, Okinawa 903-0213, Japan

Takeo Okazaki*
Department of Computer Science and Intelligent Systems
University of the Ryukyus
1 Senbaru Nishihara, Okinawa 903-0213, Japan

Abstract: The Task Technology Fit (TTF) model is a key framework in information systems research that examines the relationship between user task needs and technological capabilities. Structural Equation Modeling (SEM) and Bayesian Structural Equation Modeling (BSEM) are effective tools for analyzing the TTF model. SEM reveals complex relationships between observed and latent variables, while BSEM is particularly useful for dynamic analyses, incorporating prior information and updating the model in sequential steps. This study compared the performance of SEM, BSEM, and sequential Bayesian SEM in analyzing the TTF model, using Normal and Beta prior distributions. The Bayesian Information Criterion (BIC) assessed model fit, and the Root Mean Square Error (RMSE) evaluated coefficient accuracy. The results indicate that sequential BSEM effectively analyzes models like TTF in sequential conditions. The Beta distribution, known for its stability, is more suitable for sequential Bayesian models. This study introduces a new analytical framework to aid future research in information systems and sequential Bayesian analysis.

Keywords: Task Technology Fit (TTF); Structural Equation Modeling (SEM); Bayesian Structural Equation Modeling (BSEM); Sequential BSEM

I. INTRODUCTION

The rapid growth of information and communication technology, along with the use of analytical models to evaluate the performance of IT systems in organizations and communities, has made this practice increasingly common.

Goodhue and Thompson [1] introduced the Task Technology Fit (TTF) model, one of the most popular models in this area. The TTF model argues that technology can improve performance only when there is a strong match between the tasks users need to accomplish and the technology's capabilities. It also examines how well technology supports users' tasks and identifies the conditions under which technology boosts efficiency and productivity.

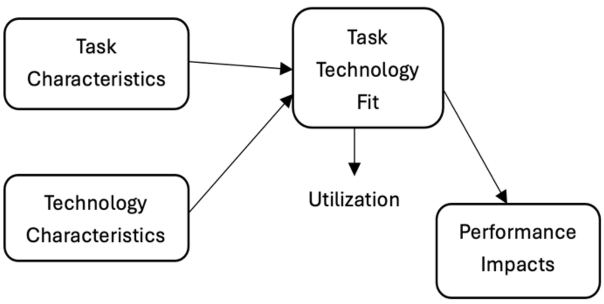

Figure 1. The Task Technology Fit Model

One of the most widely used methods to analyze TTF is Structural Equation Modeling (SEM) through mathematical models. SEM tests hypothesized patterns of directional and nondirectional relationships among a set of observed (measured) and unobserved (latent) variables [2].

The general expression for the SEM (1) is as follows:

$$\eta = B\eta + \Gamma\xi + \zeta \quad (1)$$

Where η is the vector of latent dependent variables, ξ is the vector of latent independent variables, B is the coefficient matrix representing relationships among latent dependent variables, Γ is the coefficient matrix for the effects of latent independent variables on latent dependent variables, and ζ is the vector of structural disturbances.

Recently, Bayesian Structural Equation Modeling (BSEM) has gained increased attention. The approach allows for the incorporation of prior knowledge and its update as new data becomes available. This is particularly useful when data are collected continuously over time. The Bayesian approach is well recognized in the statistics literature as an attractive approach to analyzing a wide variety of models [3][5].

The traditional equation for BSEM is based on Bayes' theorem, where prior distributions and likelihood functions are combined to estimate posterior distributions of parameters in a SEM. (2)

$$P(\Theta \mid Y) \propto P(Y \mid \Theta)\, P(\Theta) \quad (2)$$

Where $P(\Theta \mid Y)$ is the posterior distribution of the parameters (Θ) given the observed data (Y), $P(Y \mid \Theta)$ is the likelihood function, representing how the observed data are generated given the model parameters, $P(\Theta)$ is the prior distribution, encapsulating prior beliefs or information about the parameters before observing the data, and $\propto$ implies proportionality, as the equation is normalized to ensure the posterior is a valid probability distribution [4].

A key challenge in Bayesian analysis is selecting the appropriate prior distribution for the model parameters. The Normal distribution and Beta distribution are two common options. Due to its simplicity and symmetrical properties, Normal distribution is often the default choice. However, the Beta Distribution is more flexible and is better suited for modeling parameters with bounded values.

This study proposes a novel framework for analyzing the TTF model by comparing methods such as SEM, Bayesian SEM, and sequential Bayesian SEM. The data are used comprehensively and sequentially. We evaluate the framework by assessing Normal and Beta prior distributions for model fit with Bayesian Information Criterion (BIC) and parameter estimate accuracy with Root Mean Square Error (RMSE).

This research used 200 real data samples from Alessandro et al.[10]. Due to insufficient data for BSEM, ten simulated datasets of 2,000 samples each were created based on estimates from a real data SEM model to assess their similarity to the real data.

The dataset analysis involved two phases: static analysis using SEM and BSEM on 2,000 samples, and sequential analysis dividing the dataset into two parts of 1,000 samples each. The posterior distributions from the first dataset served as priors for the BSEM model on the second dataset.

This paper presents a novel approach to analyzing the TTF model, comparing Normal and Beta prior distributions in SEM and BSEM. It emphasizes the benefits of a sequential Bayesian methodology for sequential data analysis, aiming to help researchers adopt more effective tools for TTF modeling.

II. DATA CREATION

This study uses an international reference article [10] to model TTF and assess the performance of BSEM using a provided dataset of key variables.

A. REAL DATA

The reference article provides a dataset of 200 samples with key variables related to the TTF model, capturing key variables such as Task Technology Fit (TTF), Task Orientation (TO), Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Future Learning Intentions (FLI), and Behavioral Intention (BI). It includes observed variables: comp1 to comp5 for TTF, oppform1 to oppform3 for TO, ut1 to ut6 for PU, fac1 to fac6 for PEOU, int1 to int3 for BI, and fut1 to fut3 for FLI. However, this sample size is too small for BSEM, which needs a larger sample for accurate estimates and effective model evaluation, particularly in sequential data analysis or prior distribution comparisons.

B. SYNTHETIC DATA

To address the limitations of the original dataset size, we generated ten synthetic datasets, each containing 2,000 samples. This approach ensured that the datasets accurately reflected the original data while introducing variability for analytical rigor. The methodology is outlined below:

a) Synthetic datasets are generated using SEMs to define relationships between latent and observed variables.
b) Parameters for SEMs, including path coefficients, variances, and covariances, are estimated from the original data.
c) Synthetic datasets are evaluated for alignment with SEM and real data using statistical fit indices. Comparative Fit Index (CFI) assesses model fit against a baseline model. The TLI evaluates model fit while accounting for complexity. High index values show that the synthetic datasets accurately replicate the original data's structure and relationships [8].
d) Various SEM configurations are tested, and the best model is selected based on statistical fit and theoretical validity.
e) Ten synthetic datasets are created with the selected SEM to match the original data. (shown in Figure 2)

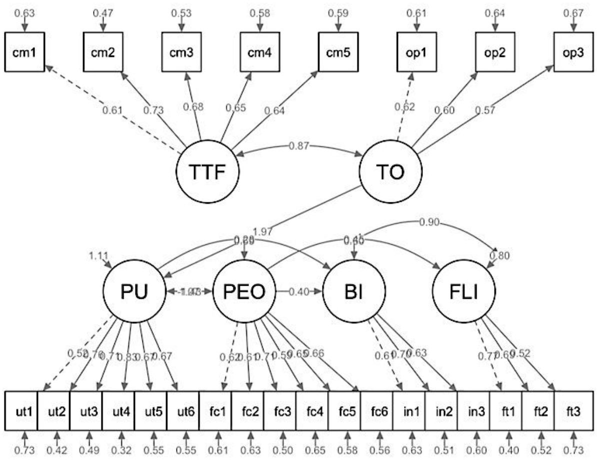

Figure 2. Selected Structural Equation Model (SEM)

The data generation process provides a strong foundation for Bayesian analyses in this study. By aligning synthetic datasets with the original data's statistical characteristics, it ensures reliability in evaluations. This generated data allows for a thorough assessment of model performance under static and sequential conditions, essential for advanced methods. It also facilitates a comparison of prior distributions, specifically

Normal and Beta, supporting meaningful insights and research advancement.

III. PRIOR DISTRIBUTIONS CHARACTERISTICS

This study focuses on analyzing the characteristics and performance of the Normal and Beta prior distributions in BSEM. Choosing an appropriate prior distribution is critical to the accuracy of coefficient estimates and model fit in Bayesian models [6]. These two distributions are also utilized as priors for analyzing the TTF model, and the results are compared. This section aims to identify each distribution's strengths and weaknesses and assess their impact on Bayesian model performance under static conditions.

A. FEATURES OF DISTRIBUTIONS

The Normal distribution, recognized for its simplicity and symmetry, is a key tool in statistical modeling, essential in many applications. The Normal distribution serves as a benchmark in this study, given its common use as a default prior in Bayesian methods, enabling robust comparisons with other prior distributions. This choice emphasizes the distribution's importance and ensures a meaningful evaluation of its effectiveness in the study.

The Beta distribution is ideal for modeling variables in the [0,1] range, making it suitable for probabilities or ratios. Its flexible shape allows for precise modeling of diverse behaviors, making it a strong choice for predicting parameters within bounded intervals and ensuring alignment with theoretical and practical needs [8].

IV. EVALUATION OF METHODS

Two indices were selected to compare the distributions. BIC is a model selection tool. This criterion evaluates model fit, where lower values indicate better accuracy and less complexity. The BIC favors models that explain the data well while minimizing unnecessary parameters [7].

RMSE is a key metric for assessing prediction accuracy. RMSE evaluates the accuracy of coefficients. A lower RMSE signifies greater accuracy in estimates.

The evaluation and comparison were based on three scenarios: The average indices across conditions were analyzed. The lowest BIC and RMSE values were analyzed. The highest index values were analyzed. This approach provides a thorough evaluation of models and distributions from multiple perspectives.

V. EXPERIMENTS

A. Comparative Evaluation of Traditional SEM and Bayesian SEM

In this section evaluated and compared the performance of traditional SEM and Bayesian SEM through static analysis. It investigated the impact of prior distributions (Normal and Beta) in BSEM compared to traditional SEM, using a dataset of 2,000 samples.

Traditional SEM: This method analyzed 2,000 sample data points without prior distributions. This analysis uses Maximum Likelihood Estimation (MLE) to calculate model coefficients.

Bayesian SEM: This method used two types of prior distributions to estimate model coefficients. A standard Normal distribution (mean zero, variance one) served as the prior for the model coefficients. The Beta distribution was defined with suitable parameters for the coefficient range, and its flexibility was used in the analysis.

1) Analysis of Results

These experiments highlight key observations on Normal and Beta priors in Bayesian modeling. The average BIC was lower for the Normal prior, indicating its better balance of model fit and complexity (Table *I*). In the worst case (the most challenging or least favorable scenario), Normal priors showed a better BIC than Beta priors (Table *II*). In the best case (the most ideal or favorable scenario), Normal priors showed the best performance, highlighting their robustness (Table *III*).

The Normal prior generally showed lower RMSE values than the Beta prior, indicating more accurate coefficient estimations. The Beta prior, while competitive, generally showed higher RMSE values, indicating lower precision in its estimations.

Table I. Compared Average of Traditional SEM and BSEM

	Traditional SEM	Normal BSEM	Beta BSEM
BIC	60320.718	59251.32	59964.91
RMSE	0.7054685	1.204484	1.217048

Table II. Compared Worst Case of Traditional SEM and BSEM

	Traditional SEM	Normal BSEM	Beta BSEM
BIC	61404.35	61989.76	62686.21
RMSE	0.7739375	1.209107	1.221524

Table III. Compared Best Case of Traditional SEM and BSEM

	Traditional SEM	Normal BSEM	Beta BSEM
BIC	59053.02	54208.67	55080.59
RMSE	0.67207	1.200621	1.215127

In summary, traditional SEM demonstrated greater accuracy in coefficient estimation than the Bayesian approach. The Normal prior consistently outperformed the Beta prior in key metrics, achieving lower BIC and RMSE values. These results highlight the reliability of the Normal prior in Bayesian modeling and the precision of the traditional SEM method.

B. Comparison of Normal and Beta Priors in Sequential BSEM

Sequential BSEM is an innovative approach to analyzing data that becomes available sequentially. This method is beneficial in cases where data are entered into the model in multiple stages, and there is a need to update the coefficients and model fit at each stage. In this research, the generated synthetic data (2000 samples) were divided into two sequential subsets, and BSEM was executed using two types of prior distributions: Normal prior and Beta prior.

Table IV. Compared Average of Normal and Beta Priors in Sequential BSEM

	Stage 1 Normal	Stage 2 Normal	Stage 1 Beta	Stage 2 Beta
BIC	29897.6185	**29642.8891**	30243.0026	29642.9019
RMSE	1.204482	**1.204737**	1.217512	1.204763

Table V. Compared Worst Case of Normal and Beta Priors in Sequential BSEM

	Stage 1 Normal	Stage 2 Normal	Stage 1 Beta	Stage 2 Beta
BIC	31236.769	**31043.611**	31552.329	31043.625
RMSE	1.208771	**1.209317**	1.221049	1.209324

Table VI. Compared Best Case of Normal and Beta Priors in Sequential BSEM

	Stage 1 Normal	Stage 2 Normal	Stage 1 Beta	Stage 2 Beta
BIC	27394.07	27083.669	27812.868	**27083.664**
RMSE	1.201072	1.19992	1.215512	**1.199917**

The main objective was to assess the performance of these distributions under sequential conditions and to compare their accuracy and stability at each stage.

1) EXPERIMENTAL DESIGN

The 2000-sample dataset was divided into two equal subsets (1000 samples each):

Stage 1: The first 1000 samples were entered into the model to estimate the coefficients and evaluate the model fit.

Stage 2: The second 1000 samples were added to the model, but in this stage, the prior information was obtained from the posterior of Stage 1 to update the coefficients and model fit.

This method also used two types of prior distributions to estimate model coefficients.

Normal Distribution: In the first stage, a standard Normal distribution (mean = 0, variance = 1) was used as the prior. In the second stage, the Normal distribution parameters were updated using the values obtained from the posterior of Stage 1.

Beta Distribution: In the first stage, Beta distribution (alpha = 1, beta = 1) was used as the prior. In the second stage, the updated coefficient values from the posterior of the first stage were used to redefine the Beta distribution.

2) Analysis of Results

Table *IV*,Table *V*, andTable *VI* show the sequential analysis of RMSE and BIC indices between Stage 1 and Stage 2 reveals insights into the behavior of Normal and Beta priors under sequential conditions.

The Normal prior's performance was nearly the same in both stages for RMSE and BIC. The lack of change indicates that the Normal prior has limited adaptability in sequential processes, hindering its ability to model evolving patterns.

The Beta prior improved significantly from Stage 1 to Stage 2, especially in RMSE, achieving greater accuracy. These results show the Beta prior's adaptability during updates, improving its effectiveness in capturing sequential changes.

The sequential results show that the Normal prior maintains stability but does not effectively leverage sequential processes for improvement. In contrast, the Beta prior demonstrates significant progress, making it more suitable for sequential Bayesian modeling, where adaptability is essential.

VI. CONCLUSION

This study compares Normal and Beta priors in Bayesian modeling, evaluating their performance in static and sequential conditions through comprehensive experiments. The findings reveal distinct strengths and limitations, offering insights for selecting priors in different contexts.

The Normal prior excelled in static settings, achieving the lowest BIC and RMSE values, and demonstrating robustness for stability-focused models. However, it showed static performance in sequential processes, indicating a lack of adaptability to evolving data structures.

The Beta prior showed strong adaptability in sequential processes, significantly improving RMSE across stages. This highlights its effectiveness in modeling changes over time. While it performed slightly worse than the Normal prior in static conditions, its ability to improve in sequential contexts makes it a valuable choice for analysis. Beta improves with more datasets, increasing its adaptability in modeling sequential processes.

In conclusion, the findings emphasize the need to align prior selection with the modeling context. The Normal prior is best for static modeling, while the Beta prior suits sequential processes. These results suggest future exploration of hybrid approaches that merge the stability of the Normal prior with the adaptability of the Beta prior, opening new avenues in Bayesian modeling.

VII. ACKNOWLEDGMENT

The authors express their sincere gratitude to Alessandro for generously providing the real data used in this study. This contribution was invaluable in enhancing the rigor and applicability of our research.

VIII. REFERENCES

[1] D. L. Goodhue and R. L. Thompson, "Task-technology fit and individual performance," Technology Studies, vol. 2, no. 1, pp. 1-19, 1995

[2] C. MacCallum and J. Austin, "Applications of Structural Equation Modeling in Psychological Research," Annual Review of Psychology, vol. 51, pp. 201-226, 2000.

[3] S.-Y. Lee, Structural Equation Modeling: A Bayesian Approach. Hoboken, NJ: Wiley, 2007.

[4] E. C. Merkle and Y. Rosseel, "blavaan: Bayesian structural equation models via parameter expansion," Journal of Statistical Software, vol. 85, no. 4, pp. 1–30, 2018. doi:10.18637/jss.v085.i04.

[5] L. van de Schoot, P. Lugtig, and J. Hox, "A tutorial on the Bayesian approach for analyzing structural equation models," International Journal of Behavioral Development, vol. 36, no. 1, pp. 84–94, 2012. doi:10.1177/0165025412436984.

[6] R. E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," Journal of the American Statistical Association, vol. 91, no. 435, pp. 1343–1370, 1996. doi:10.1080/01621459.1996.10477003.

[7] Y. Fan, J. Chen, G. Shirkey, R. John, S. R. Wu, H. Park, and C. Shao, "Applications of structural equation modeling (SEM) in ecological studies: an updated review," Ecological Processes, vol. 5, no. 1, pp. 19–30, 2016. doi:10.1186/s13717-016-0063-3.

[8] D. W. W. Ng et al., "The study of properties on generalized Beta distribution," J. Phys.: Conf. Ser., vol. 1132, p. 012080, 2018.

[9] D. Hooper, J. Coughlan, and M. R. Mullen, Structural equation modelling: Guidelines for determining model fit, Dublin Institute of Technology, Dublin, Republic of Ireland; Florida Atlantic University, Florida, USA, n.d.

[10] D. Dishaw and D. Strong, "Integrating technology acceptance and task–technology fit models for predicting behavioural and future learning intentions," Decision Support Systems, vol. 29, no. 1, pp. 9–25, 2000. doi:10.1016/S0167-9236(00)00076-2.