Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 11, No. 3, July 2026

Research Paper

Semiparametric Path Analysis with Truncated Spline: A Simulation Study with Double Resampling Inference

Fachira Haneinanda Junianto[1], Adji Achmad Rinaldo Fernandes[2], Solimun[2*], Ani Budi Astuti[2], Muhammad Hisyam Lee[3]

[1]*Department of Mathematics, Brawijaya University, Malang, East Java, 65145, Indonesia*
[2]*Department of Statistics, Brawijaya University, Malang, East Java, 65145, Indonesia*
[3]*Department of Mathematical Sciences, Universiti Teknologi Malaysia, Skudai, Johor Bahru, 81310, Malaysia*
*__Corresponding author__: Solimun@ub.ac.id

Abstract
This study proposes a semiparametric path analysis framework that integrates truncated spline modeling with resampling-based inference to capture both linear and nonlinear relationships within a unified structure. The motivation arises from the limitation of conventional path analysis, which relies on linearity assumptions that are often violated in empirical data, as indicated by the Ramsey RESET test. To address this issue, a truncated spline approach is employed to flexibly model nonlinear relationships, while statistical inference is conducted using double resampling techniques. A simulation study is performed to evaluate the performance of resampling methods under varying conditions. The results show that for a sample size of n=200 with a single nonlinear relationship, the double jackknife method provides more stable and efficient estimates compared to alternative approaches. This finding motivates its application in the empirical analysis. The empirical results, based on data from East Java, Indonesia, reveal that technology access has a significant direct effect on both financial knowledge and financial literacy. A nonlinear relationship is identified between technology access and financial literacy, characterized by a threshold effect captured through truncated spline modeling. However, the indirect effect through financial knowledge is found to be statistically insignificant. Overall, the proposed approach offers a flexible and robust framework for modeling complex causal relationships and improves inference accuracy in semiparametric path analysis.

Keywords
Semiparametric Path Analysis, Truncated Spline, Double Resampling, Financial Literacy, Nonlinear Relationship, Statistical Inference

Received: 7 March 2026, Accepted: 12 May 2026
https://doi.org/10.26554/sti.2026.11.3.1054-1067

1. INTRODUCTION

Path analysis is a statistical method used to examine causal relationships among variables within a system of interrelated structural equations. Unlike the Structural Equation Modeling (SEM) approach, which involves latent variables and measurement models, path analysis specifically focuses on the relationships between directly measured observed variables (Bollen, 1989; Grapentine, 2000; Solimun et al., 2021). This approach is widely used in various fields such as economics, finance, and the social sciences due to its ability to simultaneously identify both direct and indirect effects.

However, conventional path analysis models are generally built on the assumption of linear relationships between variables. In practice, relationships between variables are often complex and not entirely linear, so forcing a linear model can lead to model specification errors and reduce the accuracy of parameter estimates (Christodoulou-Volos and Tserkezos, 2023). Therefore, a more flexible approach is needed to capture relationship patterns that are not explicitly known.

The semiparametric approach is a relevant alternative because it combines the strengths of parametric and nonparametric models within a single modeling framework. In this approach, some relationships between variables are modeled linearly and parametrically, while other relationships whose forms are unknown are estimated using a nonparametric approach (Hastie et al., 2009; Fernandes and Solimun, 2021). In general, the semiparametric path analysis model in this study is formulated in Equation (1).

$$\bar{y} = X\bar{\beta} + \sum_{g=1}^{s-1} \bar{f}_{gi}(Y_{li}) + \bar{\varepsilon}, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\bar{y}$ represents the vector of endogenous variables, $X\bar{\beta}$ denotes the linear parametric component, $\bar{f}_{gi}(Y_{li})$ represents the nonparametric function of specific variables approximated using truncated spline, and $\bar{\varepsilon}$ is the error vector assumed to have zero mean and constant variance. This representation enables the model to simultaneously capture both linear and nonlinear relationships within the path analysis framework.

One widely used nonparametric approach for approximating such functions is the truncated spline, which has the advantage of capturing local changes in the relationship pattern while maintaining the model's interpretability (Ramsay, 1988; Maatouk et al., 2024). Furthermore, the selection of optimal knot points, for example through the Generalized Cross Validation (GCV) approach, allows the model to adaptively adjust its complexity to the data (Suriaslan et al., 2025).

In the broader semiparametric literature, alternative approaches such as Generalized Additive Models (GAM) and kernel regression have also been widely applied to capture nonlinear relationships. GAM provides substantial flexibility through smooth additive functions; however, the resulting smooth components are often difficult to interpret within a causal path analysis framework, particularly when identifying structural threshold effects (Wood, 2026). Meanwhile, kernel regression is highly sensitive to bandwidth selection and may suffer from reduced estimation stability in moderate sample sizes and multivariate settings (Vinod, 2022).

The development of path analysis based on the nonparametric truncated spline approach has been conducted in several previous studies, such as Hidayat et al. (2019); Efendi et al. (2021); Fernandes et al. (2022). These studies demonstrate that the nonparametric approach enhances the model's flexibility in capturing complex relationships among variables. However, most studies still focus on parameter estimation and have not comprehensively examined aspects of statistical inference, particularly regarding the stability and accuracy of standard error estimates.

In the context of statistical inference, resampling methods such as bootstrap and jackknife have been widely used to approximate the sampling distribution of estimators, especially when the analytical form of the distribution is difficult to obtain (Robert and Casella, 2010). Nevertheless, single resampling approaches often yield less stable standard error estimates and tend to overestimate them in complex models. To address these issues, a double resampling approach was introduced as an alternative capable of improving estimation efficiency and reducing the variability of inference results (Papalia et al., 2023).

Based on this background, this study aims to develop a semiparametric path analysis model based on truncated splines estimated using the Weighted Least Squares (WLS) method, and to integrate the double resampling approach into the statistical inference process. Additionally, a comprehensive simulation study was conducted to evaluate the model's performance under various conditions, including variations in sample size, error variance levels, and configurations of parametric and nonparametric relationships within the model structure.

The main contribution of this research lies in three methodological aspects. First, this study extends conventional path analysis into a semiparametric framework capable of simultaneously accommodating linear and nonlinear structural relationships. Second, the study explicitly integrates truncated spline approximation with double resampling inference to improve estimation stability and statistical reliability in complex semiparametric settings. Third, unlike previous studies that primarily focused on spline estimation alone, this study quantitatively compares the proposed approach with alternative semiparametric methods, including Generalized Additive Models (GAM) and kernel regression, thereby providing empirical evidence regarding its predictive performance and interpretability advantages.

2. EXPERIMENTAL SECTION

2.1 General Path Modeling

Path analysis is a statistical approach used to model causal relationships among a set of observed variables within a system of structural equations. In contrast to Structural Equation Modeling (SEM), which incorporates latent variables and measurement models, path analysis focuses exclusively on directly observed variables, allowing for a more straightforward structural representation.

Let $Y = (Y_1, Y_2, \ldots, Y_p)^\top$ denote a vector of endogenous variables and $X = (X_1, X_2, \ldots, X_k)^\top$ denote a vector of exogenous variables. The general form of a path analysis model can be expressed as a system of linear structural Equation (2).

$$Y_j = \sum_{m=1}^{p} \beta_{jm} Y_m + \sum_{l=1}^{k} \gamma_{jl} X_l + \varepsilon_j, \quad j = 1, 2, \ldots, p \tag{2}$$

where β_{jm} represents the path coefficient between endogenous variables, γ_{jl} denotes the effect of exogenous variables on endogenous variables, and ε_j is the error term associated with the j-th equation.

In matrix notation, the system can be compactly written in Equation (3).

$$Y = BY + \Gamma X + \varepsilon \tag{3}$$

where $B \in \mathbb{R}^{p\times p}$ is the matrix of endogenous coefficients, $\Gamma \in \mathbb{R}^{p\times k}$ is the matrix of exogenous coefficients, and $\varepsilon \in \mathbb{R}^p$ is the vector of random errors.

Assuming that $(I - B)$ is nonsingular, the reduced form of the model can be obtained in Equations (4).

$$Y = (I - B)^{-1}\Gamma X + (I - B)^{-1}\varepsilon \tag{4}$$

This representation shows that each endogenous variable can be expressed as a function of exogenous variables and a transformed error term, enabling the identification of both direct and indirect effects within the system.

The classical path analysis framework relies on several key assumptions, including linearity of relationships, independence of error terms, and homoscedasticity. However, these assumptions are often violated in empirical applications, particularly when the underlying relationships exhibit nonlinear patterns. This limitation motivates the extension toward a more flexible modeling framework, namely the semiparametric path model, which will be discussed in the following section.

2.2 Semiparametric Path Model

Semiparametric path analysis is an extension of the classical path analysis framework that integrates both parametric and nonparametric components within a unified modeling structure. This approach allows certain relationships among variables to be specified in a parametric (linear) form, while others with unknown functional forms are modeled nonparametrically. Such flexibility is particularly important when the linearity assumption does not hold for all relationships in the model.

In empirical applications, diagnostic tools such as the Ramsey RESET test are often used to assess model specification. The results of such tests may indicate that while some relationships can be adequately represented using linear or low-order polynomial functions, others exhibit nonlinear patterns and therefore require a more flexible, data-driven representation. This motivates the use of a semiparametric framework.

Let $Y \in \mathbb{R}^p$ denote the vector of endogenous variables and $X \in \mathbb{R}^{n\times k}$ denote the design matrix for the parametric component. The general form of the semiparametric path model can be expressed in equation (5).

$$Y = X\beta + \sum_{g=1}^{s} f_g(Z_g) + \varepsilon, \quad i = 1, 2, \dots, n \tag{5}$$

where:

- $\beta \in \mathbb{R}^k$ is the vector of parametric coefficients,
- $f_g(Z_g)$ represents the g-th nonparametric component with unknown functional form,
- Z_g is the predictor associated with the g-th nonparametric function,
- s denotes the number of nonparametric components,
- $\varepsilon \in \mathbb{R}^n$ is the error vector assumed to be independently distributed with mean zero and constant variance σ^2.

The semiparametric formulation enables the model to simultaneously capture linear and nonlinear relationships within the path structure, particularly when certain variables-such as mediators exhibit complex effects that cannot be adequately approximated using linear functions.

In this study, each nonparametric component $f_g(Z_g)$ is approximated using a truncated spline basis. The truncated spline approach is chosen due to its flexibility in capturing local variations in the data while maintaining interpretability of the model.

For illustration, consider a simple case with a single knot K and a linear spline basis. The nonparametric function can be approximated in equation (6).

$$f(Z_i) = \beta_1 Z_i + \beta_2 (Z_i - K)_+ \tag{6}$$

where the truncated spline basis function $(Z_i - K)_+$ is defined in equation (7).

$$(Z_i - K)_+ = \begin{cases} Z_i - K, & \text{if } Z_i > K \\ 0, & \text{if } Z_i \le K \end{cases} \tag{7}$$

In the context of path analysis, this formulation allows certain structural relationships-particularly those involving intermediate (mediating) variables to be modeled in a flexible nonlinear form, while other relationships remain linear. As a result, the semiparametric path model provides a more accurate representation of complex causal mechanisms compared to conventional linear path models.

2.3 Truncated Spline Approximation

To estimate the unknown nonparametric components in the semiparametric path model, this study employs a truncated spline approach. Truncated splines are widely used due to their flexibility in approximating nonlinear relationships while preserving interpretability and computational efficiency.

In general, a truncated spline function of degree m with r knots can be expressed as equation (8).

$$f(Z_i) = \beta_0 + \sum_{j=1}^{m} \beta_j Z_i^j + \sum_{k=1}^{r} \delta_k (Z_i - K_k)_+^m \tag{8}$$

where:

- $\beta_0, \beta_1, \dots, \beta_m$ are polynomial coefficients,
- δ_k are coefficients associated with the truncated components,
- K_k denotes the k-th knot location,
- $(Z_i - K_k)_+^m$ is the truncated power basis function defined as Equation (9).

$$(Z_i - K_k)_+^m = \begin{cases} (Z_i - K_k)^m, & \text{if } Z_i > K_k \\ 0, & \text{if } Z_i \le K_k \end{cases} \tag{9}$$

This formulation allows the regression function to change its behavior at different regions of the predictor space, enabling the model to capture local nonlinear patterns.

For notational convenience, the truncated spline function can be rewritten in a linear-in-parameter form as Equation (10).

$$f(Z_i) = b(Z_i)^\top \theta \tag{10}$$

where:

- $b(Z_i)$ is the vector of spline basis functions,

- θ is the corresponding parameter vector.

Specifically, the basis function vector can be written as Equation (11).

$$b(Z_i) = \left(1, Z_i, Z_i^2, \ldots, Z_i^m, (Z_i - K_1)_+^m, \ldots, (Z_i - K_r)_+^m\right)^\top \tag{11}$$

and the parameter vector is θ. By stacking all observations, the nonparametric component can be expressed in matrix form as Equation (12).

$$f = B\theta \tag{12}$$

where:

- $B \in \mathbb{R}^{n\times(m+r+1)}$ is the spline basis matrix,
- each row of B corresponds to $b(Z_i)^\top$.

Substituting this representation into the semiparametric path model, the overall model can be written as Equation (13).

$$Y = X\beta + B\theta + \varepsilon \tag{13}$$

This formulation shows that the semiparametric model can be transformed into a linear regression model with respect to the unknown parameters β and θ. Therefore, standard estimation techniques such as Weighted Least Squares (WLS) can be applied.

Furthermore, when multiple nonparametric components are involved, the spline basis matrices corresponding to each component can be concatenated into a single augmented matrix, resulting in a unified representation in Equation (14).

$$Y = X\beta + B^*\theta^* + \varepsilon \tag{14}$$

where B^* and θ^* denote the combined spline basis matrix and parameter vector, respectively.

This matrix-based representation plays a crucial role in simplifying the estimation procedure and provides a foundation for the development of efficient inference methods in the subsequent section.

2.4 Double Resampling Inference

Resampling is a widely used statistical approach for obtaining improved estimates of population parameters or sampling distributions by repeatedly drawing samples from the observed data. In general, resampling methods generate new samples from the original dataset either with replacement, as in bootstrap, or without replacement, as in jackknife. These approaches are particularly useful when analytical derivation of sampling distributions is difficult or when classical assumptions such as normality are not satisfied (Hastie et al., 2009; Amanda et al., 2024).

While single resampling methods provide a practical way to estimate variability, they may still produce unstable or biased standard error estimates in complex models. To address this limitation, double resampling is introduced as an extension that incorporates an additional layer of resampling, allowing for a more comprehensive assessment of uncertainty (Hastie et al., 2009). In this study, two types of double resampling are employed, namely double bootstrap and double jackknife, to improve the accuracy and stability of inference in the semiparametric path model.

In the double bootstrap procedure, resampling is performed in two stages. First, B_1 bootstrap samples are generated from the original dataset with replacement. Then, for each first-stage bootstrap sample, a second-stage bootstrap is conducted by generating B_2 additional resamples. As a result, the total number of computed statistics becomes $B_1 + B_1B_2$ (Papalia et al., 2023 MacKinnon, 2006). Let $\hat{\beta}^{*(r)}$ denote the parameter estimate obtained from the r-th second-stage bootstrap sample, and let $\bar{\beta}^*$ denote the average of these estimates. The standard error of the double bootstrap is given by Equation (15).

$$\widehat{(se)}_{BB} = \sqrt{\frac{1}{B_2}\sum_{r=1}^{B_2}\left(\hat{\beta}^{*(r)} - \bar{\beta}^*\right)^2} \tag{15}$$

Alternatively, the double jackknife procedure is based on systematic deletion of observations without replacement. In contrast to bootstrap, which samples with replacement, jackknife repeatedly removes observations from the dataset to assess variability (Asare et al., 2024). In the first stage, each observation is deleted one at a time from a sample of size n, producing n subsamples of size $n-1$. For each subsample, a second-stage jackknife is applied, resulting in a nested resampling structure. Let $\hat{\beta}_{(r)}$ denote the parameter estimate obtained from the r-th double jackknife sample. The standard error is computed as Equation (16).

$$\widehat{(se)}_{JJ} = \sqrt{\frac{(n-d)}{d(n-d)}\sum_{r=1}^{n}\left(\hat{\beta}_{(r)} - \bar{\hat{\beta}}\right)^2} \tag{16}$$

where d denotes the number of deleted observations.

For both approaches, statistical inference on the path coefficients is conducted using the following test statistic in Equation (17).

$$t_{\text{obs}} = \frac{\hat{\beta}_{(.)}}{\widehat{se}(\hat{\beta}_{(.)})} \tag{17}$$

with hypotheses $H_0 : \beta_r = 0$, $H_1 : \beta_r \neq 0$. The null hypothesis is rejected if the corresponding p-value is less than the significance level α, indicating a statistically significant relationship between variables.

In the context of the proposed semiparametric path model, double resampling plays a crucial role in improving inference quality. Since the model combines parametric and nonparametric components and is estimated using WLS, deriving analytical standard errors can be challenging. Therefore, double bootstrap and double jackknife are utilized to obtain more stable standard error estimates and more reliable hypothesis testing results. Compared to single resampling methods, double resampling is expected to better capture sampling variability by incorporating an additional layer of uncertainty, making it particularly suitable for semiparametric models involving truncated spline components.

2.5 Simulation Study

To evaluate the performance of the proposed semiparametric path model with truncated spline and double resampling inference, a comprehensive simulation study is conducted under various experimental settings. The simulation is designed to assess both the accuracy of parameter estimation and the reliability of inferential procedures (Junianto et al., 2025).

The data-generating process (DGP) follows a semiparametric path structure that combines parametric and nonparametric relationships, defined as Equation (18).

$$\begin{aligned} Y_1 &= \beta_{10} + \beta_{11}X + \varepsilon_1, \\ Y_2 &= \beta_{20} + \beta_{21}Y_1 + g(X) + \varepsilon_2, \end{aligned} \tag{18}$$

where $g(X)$ represents a nonlinear function approximated using truncated spline basis functions. The error terms ε_1 and ε_2 are generated independently from a normal distribution with mean zero and variance σ^2. The true parameter values are fixed across all simulation scenarios to ensure consistent evaluation of estimator performance.

Several simulation scenarios are constructed by varying key factors, including sample size, error variance, and structural configuration of the model. The sample sizes are defined as $n \in \{25, 200, 1000\}$, representing small, moderate, and large sample conditions. Different levels of error variance are considered to examine the robustness of estimation and inference under varying noise conditions. For each scenario, the simulation is replicated R times to ensure stability and reliability of the results.

In addition, the structural configuration of the model is varied through six scenarios that represent different combinations of linear and nonparametric relationships among variables. These scenarios are summarized in Table 1.

These scenarios are designed to capture various possible configurations of nonlinear effects within the path structure, allowing a comprehensive evaluation of the proposed semiparametric modeling approach.

For each generated dataset, parameter estimation is performed using the Weighted Least Squares (WLS) method based on the linearized semiparametric path model. Statistical inference is conducted using four resampling approaches, namely bootstrap (B), jackknife (J), double bootstrap (BB), and double jackknife (JJ). Each resampling method is applied consistently across all simulation scenarios to allow a fair comparison of their performance (James et al., 2023). Furthermore, all simulation scenarios were conducted using 1,000 replications to ensure the stability and reliability of the estimation and inference results.

Table 1. Structural Relationship Scenarios in the Simulation Study

Scenario	$(X_1 \to Y_1)$	$(X_1 \to Y_2)$	$(Y_1 \to Y_2)$
1	Linear	Nonparametric	Linear
2	Linear	Linear	Nonparametric
3	Nonparametric	Linear	Linear
4	Linear	Nonparametric	Nonparametric
5	Nonparametric	Linear	Nonparametric
6	Nonparametric	Nonparametric	Linear

To improve the comprehensiveness of the simulation design, additional robustness scenarios were incorporated into the study. Besides the standard normal distribution, the error terms were also generated from heavy-tailed and skewed distributions. Heavy-tailed errors were simulated using a Student-*t* distribution with low degrees of freedom, while skewed errors were generated using a log-normal distribution. Furthermore, an outlier contamination scenario was introduced by replacing 5% of the observations with extreme values to evaluate the robustness of the estimators under data irregularities.

The performance of the estimators is evaluated using three main criteria. First, the bias of the estimator is defined as Equation (19).

$$\mathrm{Bias}(\hat{\beta}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\beta}_r - \beta) \tag{19}$$

which measures the average deviation of the estimated parameter from its true value. Second, the standard error (SE) is computed as Equation (20).

$$\mathrm{SE}(\hat{\beta}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\beta}_r - \bar{\hat{\beta}})^2} \tag{20}$$

where $\bar{\hat{\beta}}$ denotes the mean of the parameter estimates across all replications. Third, the relative efficiency (RE) is defined as Equation (21).

$$\mathrm{RE} = \frac{\mathrm{SE}_{\mathrm{method}}}{\mathrm{SE}_{\mathrm{benchmark}}} \tag{21}$$

where the benchmark is selected as one of the resampling methods, such as the bootstrap approach, to facilitate comparative evaluation.

2.6 Data Description

The empirical analysis in this study utilizes primary data obtained through a questionnaire-based survey. The dataset has undergone prior validity and reliability testing, ensuring that the measurement instruments are statistically sound and suitable for further analysis.

The population of this study consists of adult individuals residing in East Java, who are considered potential bank customers. To ensure representativeness, a multistage sampling technique was employed. In the first stage, districts and cities were selected using judgment sampling based on three major cultural regions in East Java, namely Arek, Mataraman, and Madura. In the second stage, sub-districts or villages were selected from both urban and suburban areas within each selected region. In the final stage, respondents were selected using a combination of quota sampling and accidental sampling.

The accidental sampling approach was applied by selecting individuals who met the predefined criteria, including a minimum age of 20 years, at least a high school level of education, current employment status, and a minimum residency duration of three years in the selected area. Data collection was conducted using structured questionnaires distributed directly to respondents.

The total sample size in this study is $n = 200$ respondents, obtained through a multistage sampling procedure combining judgment sampling, quota sampling, and accidental sampling techniques. This sample size satisfies the recommended criteria for structural and semiparametric modeling. According to Hair et al. (2019), an adequate sample size should follow a minimum ratio of 5:1 to 20:1 relative to the number of estimated parameters. Given the complexity of the semiparametric path model, the sample size used in this study is considered statistically sufficient to ensure stable parameter estimation, reliable inference, and robust model performance.

This study adopts a data-driven approach, where the observed data are analyzed without imposing strong parametric assumptions regarding the underlying relationships. Unlike previous studies that employ Bayesian estimation, this study focuses on a frequentist framework using semiparametric modeling combined with resampling-based inference, allowing the data to reveal both linear and nonlinear patterns.

Three main variables are considered in this study:

- X: Technology Access, representing the level of access to financial and digital technology,
- Y_1: Financial Knowledge, representing individuals' understanding of financial concepts,
- Y_2: Financial Literacy, representing the overall financial capability of individuals.

In the context of path analysis, Technology Access (X) is treated as an exogenous variable, financial knowledge (Y_1) as an intermediate (mediating) variable, and financial literacy (Y_2) as the final endogenous variable. This structure enables the analysis of both direct and indirect effects, as well as the potential presence of nonlinear relationships within the model.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics and Variable Measurement

This study uses three main variables technology access (X_1), financial knowledge (Y_1), and financial literacy (Y_2). All variables are measured using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), which is widely used in behavioral and economic research to quantitatively capture individuals' perceptions and levels of understanding.

Operationally, technology access (X_1) represents the level of accessibility and utilization of digital financial technology by individuals, measured through indicators of accessibility to technology-based financial services, intensity of use, and availability of supporting infrastructure. The financial knowledge variable (Y_1) reflects an individual's cognitive capacity to understand basic financial concepts, measured through indicators of understanding of interest rates, inflation, and risk diversification as fundamental components in financial decision-making. Meanwhile, financial literacy (Y_2) is positioned as the primary outcome, representing an individual's ability to integrate financial knowledge into rational and effective financial behavior and decisions.

The measurement of variables in this study is based on a reflective approach, where the indicators used are assumed to be manifestations of the latent constructs being measured. Therefore, prior to conducting structural analysis, an initial evaluation of data characteristics was performed through descriptive statistical analysis and correlation analysis. This analysis aims to identify data distribution, the degree of variation, and patterns of relationships among variables as a basis for ensuring model suitability and avoiding potential statistical issues such as multicollinearity.

Table 2. Descriptive Statistics and Correlation Matrix

Variable	Mean	Std. Dev	X_1	Y_1	Y_2
X_1	3.85	0.72	1		
Y_1	3.67	0.68	0.54	1	
Y_2	3.72	0.70	0.49	0.58	1

The results of the descriptive and correlation statistical analyses are presented in Table 2 as a basis for initial exploration of the data structure prior to further semiparametric modeling.

Based on the table, all variables have mean values above the midpoint of the Likert scale (> 3), indicating that respondents generally have relatively high levels of technology access, financial knowledge, and financial literacy. This suggests that the majority of respondents have had sufficient exposure to financial technology and possess a basic understanding of financial concepts, thereby conceptually supporting the development of better financial literacy.

Furthermore, standard deviation values ranging from 0.68 to 0.72 indicate a moderate level of data dispersion. This indicates that although there is variation among respondents, the data distribution remains within a stable range and does not show any dominant outliers. Statistically, this condition is important because it ensures that parameter estimates in the model are not dominated by excessive variation, thereby enhancing the reliability of the analysis results.

The correlation matrix shows that all variables have positive relationships with correlation strengths falling into the moderate category, ranging from 0.49 to 0.58. These relationships indicate a structural interdependence consistent with the theoretical framework, where an increase in Technology Access tends to be followed by an increase in Financial Knowledge and Financial Literacy. This pattern provides initial justification for the appropriateness of using a path analysis model, as the relationships among variables meet the assumption of linear interdependence as the basis for establishing a causal structure.

Furthermore, no very high correlation coefficients (≥ 0.90) were found, so it can be concluded that there is no indication of serious multicollinearity among the independent variables in the model. This condition is crucial for ensuring the stability of parameter estimates, as high multicollinearity can lead to increased estimation variance and reduce the accuracy of interpreting path coefficients.

3.2 Classical Assumptions

Before estimating the semiparametric model, tests of classical assumptions were conducted to ensure that the model used satisfies the basic assumptions within the framework of regression and path analysis. These tests aim to ensure that the resulting parameter estimates are unbiased, consistent, and efficient, particularly in the context of a model that combines parametric and nonparametric components.

The tests conducted include multicollinearity, heteroscedasticity, and residual normality. Given the model structure consisting of several equations, the tests focused on the structural equation with financial literacy (Y_2) as the dependent variable, and technology access (X_1) and financial knowledge (Y_1) as independent variables. The test results are presented in Table 3.

Based on Table 3, the results of the multicollinearity test show that the Variance Inflation Factor (VIF) values for the variables technology access (X_1) and financial knowledge (Y_1) are 2.35 and 2.78, respectively, which are below the critical threshold of 10. This indicates that there is no high linear relationship between the independent variables, so the model is free from multicollinearity issues and parameter estimation can be performed stably.

Furthermore, the results of the heteroscedasticity test using the Breusch–Pagan method yielded a p-value of 0.214, which is greater than 0.05. Thus, there is insufficient evidence to reject the null hypothesis, so it can be concluded that the residual variance is constant (homoscedastic).

Furthermore, the results of the residual normality test using the Shapiro-Wilk test show a p-value of 0.087, which is also greater than the significance level of 0.05. This indicates that the residuals are normally distributed, thus fulfilling an important assumption in statistical inference.

3.3 Linearity Test Using Ramsey's RESET

The validity of path analysis critically depends on the linearity assumption among variables. Therefore, testing linearity constitutes an essential preliminary step before parameter estimation. In this study, linearity is examined using the Ramsey' RESET (Regression Equation Specification Error Test) and its modified version, which allows for the detection of both parametric and nonparametric nonlinear relationships.

The baseline model used in this study is specified as a linear regression of the form equation (22).

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{22}$$

To test for potential nonlinearities, the model is augmented by incorporating higher-order terms of the independent variable in equation (23).

$$Y = \beta_0 + \beta_1 X + \gamma_1 X^2 + \gamma_2 X^3 + \varepsilon \tag{23}$$

The null hypothesis states that the linear specification is adequate (i.e., $\gamma_1 = \gamma_2 = 0$). Rejection of the null hypothesis indicates the presence of nonlinear effects that cannot be captured by a purely linear model. The results of the linearity tests are presented in Table 4.

Based on Table 4, the relationships between technology access (X_1) and Financial Knowledge (Y_1), as well as between financial knowledge (Y_1) and financial literacy (Y_2), are found to be linear, with p-values of 0.9667 and 0.9979, respectively. Since both p-values exceed the significance level of 0.05, the null hypothesis of linearity is not rejected, indicating that these relationships can be adequately modeled using a parametric linear specification.

In contrast, the relationship between technology access (X_1) and financial literacy (Y_2) exhibits a nonlinear pattern, as indicated by a p-value of 0.0348 obtained from the Ramsey RESET test. Given that this value is less than 0.05, the null hypothesis of linearity is rejected, suggesting model misspecification under a purely linear assumption. Consequently, further analysis is conducted using the modified Ramsey RESET test to explore potential nonlinear structures.

The results of the modified Ramsey RESET indicate that parametric nonlinear forms, such as quadratic and cubic models, yield p-values greater than 0.05, implying that these specifications are insufficient to capture the underlying relationship. However, the nonparametric approach using truncated spline demonstrates statistically significant results, with p-values of 0.0005 for the linear spline with one knot and 0.0023 for the quadratic spline with one knot.

To determine the most appropriate model, statistical power considerations are taken into account. Lower p-values are associated with higher test power, indicating a stronger ability to detect true nonlinear relationships. Based on this criterion, the truncated spline model with a linear basis and one knot is selected as the optimal specification for modeling the relationship between technology access and financial literacy.

Table 3. Classical Assumption Diagnostics

Assumptions	Method	Statistic Value	Threshold
Multicollinearity	VIF (X_1)	2.35	VIF < 10
	VIF (Y_1)	2.78	VIF < 10
Heteroscedasticity	Breusch–Pagan Test	p-value = 0.214	$p > 0.05$
Normality	Shapiro–Wilk Test	p-value = 0.087	$p > 0.05$

Table 4. Ramsey's Results

Relationship Between Variables	Ramsey's RESET (p-value)	Parametric (Quadratic)	Truncated Spline (Linear Knot)	Truncated Spline (Quadratic Knot)	Conclusion
($X_1 \to Y_1$)	0.9667	–	–	–	Linear
($X_1 \to Y_2$)	0.0348	0.5626	0.0005	0.0023	Nonlinear (Spline)
($Y_1 \to Y_2$)	0.9979	–	–	–	Linear

These findings suggest that the effect of technology access on financial literacy is not uniform across all levels of technology access. Instead, the presence of a knot indicates a structural change in the relationship, where the marginal effect of technology access varies depending on the level of the predictor.

From a practical perspective, this implies that policies aimed at improving financial literacy through technology access should not adopt a uniform strategy. Instead, targeted interventions that consider threshold effects in technology adoption may yield more effective outcomes. Therefore, the use of a semiparametric truncated spline model is not only methodologically justified but also provides meaningful insights for data-driven decision-making.

3.4 Model Specification

Based on the linearity testing results using the Ramsey RESET test and its modified version, the structural relationships in the model can be classified into linear and nonlinear components. Specifically, the relationships between technology access (X_1) and financial knowledge (Y_1), as well as between financial knowledge (Y_1) and financial literacy (Y_2), are identified as linear. In contrast, the relationship between technology access (X_1) and financial literacy (Y_2) is found to be nonlinear and is therefore modeled using a truncated spline approach with one knot.

Furthermore, the empirical data structure characterized by a sample size of $n = 200$ and a single nonlinear relationship is consistent with the simulation scenario that demonstrated superior performance of the double jackknife method. Therefore, the semiparametric path model is specified accordingly.

The optimal knot location is primarily determined using the Generalized Cross Validation (GCV) criterion, where the best model corresponds to the minimum GCV value. GCV was selected as the main criterion because it effectively balances model fit and model complexity in semiparametric truncated spline estimation.

Figure 1 presents the relationship between the GCV values and the candidate knot locations. Based on Figure 1, the optimal knot is identified at K=3.667, corresponding to the minimum GCV value. This indicates that the truncated spline model with one knot at this location provides the most appropriate balance between flexibility and parsimony.

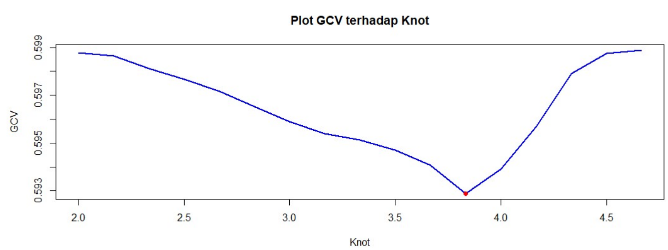

Figure 1. Plot of GCV Versus Knot

To further evaluate the robustness of the knot selection procedure, a sensitivity analysis was conducted using several alternative evaluation criteria, namely Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and 5-Fold Cross-Validation (CV). Several candidate knot locations were assessed under all criteria, and the comparison results are presented in Table 5.

Table 5. Sensitivity Analysis of Knot Selection

Knot Location	GCV	AIC	BIC	5-Fold CV
3.200	0.287	142.61	149.22	0.294
3.400	0.241	134.73	141.08	0.248
3.667	0.214	128.53	135.72	0.219
3.800	0.226	130.47	137.95	0.231
4.000	0.259	136.18	143.66	0.267

Table 5 shows that the knot location at K=3.667 consistently yields the smallest values of GCV, AIC, BIC, and 5-fold CV among all evaluated candidates. These findings indicate that the selected knot location offers the best balance between model fit and model complexity. Furthermore, the consistency of results across different evaluation criteria confirms the robustness and stability of the proposed semiparametric truncated spline model.

3.5 Simulation Results

To evaluate the effectiveness of the proposed modeling framework, a comparative analysis was conducted between parametric, nonparametric, and semiparametric approaches, including GAM and kernel regression. This comparison aims to assess the ability of each model to capture the underlying data structure, particularly in the presence of nonlinear relationships identified in the previous analysis.

The evaluation was based on three key performance metrics, namely the coefficient of determination (R^2), Root Mean Square Error (RMSE), and estimation bias. The R^2 measures the explanatory power of the model, while RMSE and bias reflect the predictive accuracy and stability of the estimates. A model with higher R^2 and lower RMSE and bias is considered to provide better overall performance. The comparison results are presented in Table 6.

As shown in Table 6, the proposed semiparametric truncated spline model consistently outperforms other alternative approaches on all evaluation metrics. The parametric linear model exhibits the lowest explanatory power and the highest prediction error, indicating its inability to adequately capture the nonlinear structures present in the data. Although the quadratic specification slightly improves model fit, it remains insufficient to represent more complex nonlinear patterns.

Among the semiparametric alternatives, GAM demonstrates relatively strong predictive performance thanks to its flexible smoothing functions. However, its smoothing-function-based representation does not explicitly identify structural threshold effects, thereby limiting interpretability within a path analysis framework. Similarly, kernel regression enhances flexibility but exhibits relatively higher estimation variability due to sensitivity in window width selection.

The semiparametric truncated spline model achieves the best overall performance, yielding the highest coefficient of determination ($R^2 = 0.79$), the lowest RMSE (0.47), the highest CP (0.95), and the smallest bias (0.08). These results indicate that the semiparametric truncated spline approach provides a more effective balance between flexibility, interpretability, and estimation stability, particularly in modeling threshold-based nonlinear relationships within a semiparametric path structure.

The effect of sample size on the performance of the semiparametric path model is presented in Table 7. The results summarize the behavior of the estimator in terms of bias, standard error, relative dispersion, and model fit across different sample sizes.

As shown in Table 7, the average bias values remain relatively stable across all sample sizes, with values of 0.14908 for $n = 25$, 0.14910 for $n = 200$, and 0.14910 for $n = 1000$. This indicates that the estimator exhibits consistent bias behavior regardless of the number of observations. Although the bias does not decrease substantially as the sample size increases, its stability suggests that the estimation process does not introduce additional systematic error under different sample conditions.

In contrast, the standard error (SE) decreases as the sample size increases. For a small sample ($n = 25$), the SE is 0.04657, which declines to 0.03419 for $n = 200$ and remains stable at 0.03422 for $n = 1000$. This pattern is consistent with statistical theory, where larger sample sizes reduce sampling variability and improve estimation precision.

The ratio SE remains relatively stable around 3 across all sample sizes, indicating that the relative dispersion of the estimator is consistent and not sensitive to sample size variation. This stability reflects the robustness of the estimator in maintaining proportional variability across different data scales.

Furthermore, the coefficient of determination (R^2) shows minimal variation, ranging from 0.71987 to 0.72075. This suggests that the explanatory power of the semiparametric path model is not significantly influenced by sample size. The model consistently captures the underlying relationship between variables, even in smaller samples.

The effect of resampling methods on the estimation performance of the semiparametric path analysis model is presented in Table 8. A comparison was made between single and multiple resampling methods to evaluate differences in the accuracy and stability of the estimates.

Based on Table 8, the choice of resampling method influences the estimation accuracy and inferential performance of the semiparametric path model. The single bootstrap and single jackknife approaches produce relatively higher average bias values, namely 0.1574 and 0.1573, respectively. In addition, both methods exhibit larger Ratio SE values, indicating greater estimation dispersion compared to the double resampling approaches.

The double bootstrap and double jackknife methods demonstrate improved estimation performance, as reflected by lower average bias values of 0.1409 and 0.1408, respectively. Furthermore, the RMSE values decrease from 0.253 and 0.247 under single resampling to 0.221 and 0.214 under double resampling, indicating higher estimation accuracy and improved predictive stability.

The Coverage Probability also increases under double resampling approaches. The double jackknife method achieves the highest Coverage Probability of 0.95, followed by double bootstrap with 0.94, suggesting that the resulting confidence intervals provide more reliable inferential performance. Meanwhile, the coefficient of determination (R^2) remains relatively stable across all resampling methods, with values around 0.72, indicating that differences among methods primarily affect estimation stability and inferential reliability rather than explanatory power.

Among the evaluated approaches, the double jackknife method demonstrates the best performance due to its lowest RMSE, smallest average bias, and highest Coverage Probability. These findings indicate that incorporating a second resampling layer improves the stability and reliability of statistical inference in semiparametric path estimation.

Table 6. Comparison of Model Performance

Model Type	Functional Form	R^2	RMSE	Average Bias	Coverage Probability
Parametric	Linear	0.68	0.62	0.15	0.86
Parametric	Quadratic	0.71	0.58	0.13	0.88
Nonparametric	Kernel Regression	0.72	0.56	0.12	0.89
Semiparametric	GAM	0.75	0.51	0.10	0.92
Nonparametric	Truncated Spline (1 Knot)	0.74	0.52	0.10	0.91
Semiparametric	Truncated Spline (1 Knot)	0.79	0.47	0.08	0.95

Table 7. Simulation Results on Sample Size

Sample Size	Average Bias	SE	Ratio SE	R^2
25	0.14908	0.04657	299.739	0.71987
200	0.14910	0.03419	299.989	0.72075
1000	0.14910	0.03422	300.236	0.72035

The effect of variations in model scenarios on estimation performance is presented in Table 9. The scenarios tested represent differences in model structure configurations, particularly regarding the position and characteristics of nonlinear components in the semiparametric path analysis model.

Based on Table 9, it can be observed that the average bias and standard error values remain relatively stable across all scenarios, with bias values around 0.149 and standard errors ranging from 0.03830 to 0.03838. This indicates that variations in model structure do not have a significant effect on the accuracy of parameter estimation.

Similarly, the ratio SE values fall within a very narrow range around 3, indicating that the relative dispersion of the estimator remains consistent across different model configurations. This consistency suggests that the estimation method exhibits strong stability against changes in the structural specification of the model.

However, a notable difference is observed in the coefficient of determination (R^2). Scenario 1 yields the highest R^2 value of 0.79480, followed by Scenario 2 with a value of 0.76094. In contrast, the remaining scenarios show lower R^2 values, approximately around 0.69.

These differences in R^2 indicate that model configuration, particularly the placement of the nonlinear component, influences the model's ability to explain data variability. Scenarios with higher R^2 values suggest that the model is better able to capture the underlying relationships among variables.

Substantively, these findings imply that the placement of the nonlinear component within the model structure plays a crucial role in improving model performance, not in terms of parameter estimation accuracy (bias and standard error), but in terms of predictive capability and overall goodness-of-fit. In other words, an appropriate specification of the semiparametric model can enhance the quality of representing relationships among variables.

The effect of variations in error variance on the model's estimation performance is presented in Table 10. This analysis aims to evaluate the robustness of the semiparametric model to changes in the level of noise in the data.

Based on Table 11, the proposed semiparametric truncated spline model maintains relatively stable estimation performance across different error distributions and outlier scenarios. Under the normal error distribution, the model achieves an RMSE of 0.47, bias of 0.08, and R^2 of 0.79. When heavy-tailed and skewed error distributions are introduced, the RMSE slightly increases to 0.51 and 0.53, respectively, while the bias rises moderately to 0.10 and 0.11. Similarly, the coefficient of determination decreases slightly but remains above 0.70. In the outlier contamination scenario, the model still demonstrates acceptable performance with RMSE of 0.56, bias of 0.13, and R^2 of 0.72. Overall, these results indicate that the proposed estimation framework is reasonably robust against deviations from classical normality assumptions and remains reliable under more realistic data conditions.

Based on the results of the simulation study, it is observed that the performance of resampling methods varies depending on the sample size and the structural characteristics of the model. In particular, for a moderate sample size of n=200 and a model configuration involving one nonlinear relationship, the double jackknife method consistently provides more stable and efficient estimates compared to other resampling approaches. This is reflected in its lower standard error and improved estimation accuracy across simulation scenarios. Therefore, considering the similarity between the simulation setting and the empirical data used in this study, the double jackknife approach is selected as the primary inference method in the subsequent empirical analysis.

3.6 Empirical Results

Based on the simulation results, the double jackknife method under Scenario 1 demonstrates the most stable and efficient estimation performance, as indicated by lower RMSE values, higher coverage probability, and more reliable inferential stability compared to other resampling approaches. Since the empirical structure of this study shares similar characteristics with Scenario 1, particularly the presence of a single nonlinear relationship within the model, the double jackknife method was selected as the primary inferential approach for the empirical analysis.

Table 8. Simulation Results on Resampling Type

Resampling Method	Average Bias	Ratio SE	R^2	RMSE	Coverage Probability
Single Bootstrap	0.1574	35.009	0.7201	0.253	0.91
Single Jackknife	0.1573	34.994	0.7204	0.247	0.92
Double Bootstrap	0.1409	24.961	0.7199	0.221	0.94
Double Jackknife	0.1408	25.031	0.7208	0.214	0.95

Table 9. Simulation Results Model Scenarios

Scenario	Average Bias	SE	Ratio SE	R^2
Scenario 1	0.14911	0.03834	299.573	0.79480
Scenario 2	0.14901	0.03830	300.111	0.76094
Scenario 3	0.14913	0.03833	299.910	0.69173
Scenario 4	0.14911	0.03838	300.625	0.69194
Scenario 5	0.14905	0.03830	299.676	0.69065
Scenario 6	0.14915	0.03830	300.032	0.69185

Table 10. Simulation Results on Error Variance Levels

Error Variance (EV)	Average Bias	SE	Ratio SE	R^2
EV = 0.5 × MSE	0.0639	0.0383	30.007	0.6977
EV = 1.0 × MSE	0.1278	0.0383	29.994	0.7663
EV = 2.0 × MSE	0.2556	0.0383	29.995	0.6969

Table 11. Evaluation Under Different Error Distributions

Error Distribution	RMSE	Average Bias	R^2
Normal	0.47	0.08	0.79
Student-t	0.51	0.10	0.76
Log-normal	0.53	0.11	0.74
5% Outliers	0.56	0.13	0.72

Based on Table 12, technology access (X_1) has a positive and statistically significant direct effect on financial knowledge (Y_1), with an estimated coefficient of 0.143 ($p < 0.001$). This result indicates that greater access to technology and digital financial services contributes to improving individuals' understanding of financial concepts. Higher levels of technology access increase individuals' opportunities to obtain financial information and gain practical experience related to digital financial activities.

Furthermore, financial knowledge (Y_1) also exerts a positive and significant effect on financial literacy (Y_2), with an estimated coefficient of 0.027 ($p < 0.001$). Although statistically significant, the magnitude of this effect is relatively smaller compared to the direct effect of technology access on financial literacy. This finding suggests that financial literacy is not solely determined by formal financial knowledge, but is also directly influenced by individuals' interaction with financial technology.

A major finding of this study is the presence of a nonlinear relationship between technology access and financial literacy, modeled using a truncated spline approach. The estimation results show that under the regime ($X_1 \leq K_1$), the effect of technology access on financial literacy is 0.743, whereas under the regime ($X_1 > K_1$), the effect decreases to 0.420. Both effects are statistically significant at the 5% significance level.

These findings indicate the existence of a threshold effect or structural change in the relationship between the variables. At lower levels of technology access, increases in access generate substantial improvements in financial literacy. However, after surpassing a certain knot point, additional increases in technology access produce relatively smaller improvements in financial literacy. This pattern reflects a diminishing marginal effect, implying that the incremental benefit of technology access declines once individuals reach a certain level of technological exposure.

From a methodological perspective, these results reinforce the importance of employing a semiparametric approach rather than relying solely on conventional linear models. If the relationship were forced into a purely linear specification, the behavioral changes across different regimes would not be adequately captured. Therefore, the truncated spline approach provides greater flexibility in modeling complex structural relationships.

Meanwhile, the indirect effect of technology access on financial literacy through financial knowledge yields an estimated value of 0.004 with a p-value of 0.307, indicating statistical insignificance. This finding suggests that financial knowledge does not function as a strong mediating variable in the relationship between technology access and financial literacy. Consequently, the influence of technology access on financial literacy occurs predominantly through direct pathways rather than through

Table 12. Results of Semiparametric Path Analysis with Double Resampling Jackknife

Path	Coefficient Estimate	p-value
Technology Access (X_1) → Financial Knowledge (Y_1)	0.143	<0.001
Technology Access (X_1) → Financial Literacy (Y_2) for $X_1 \leq K_1$	0.743	<0.001
Technology Access (X_1) → Financial Literacy (Y_2) for $X_1 \leq K_2$	0.420	<0.001
Financial Knowledge (Y_1) → Financial Literacy (Y_2)	0.027	<0.001
Technology Access (X_1) → Financial Literacy (Y_2) through Financial Knowledge (Y_1)	0.004	0.307

the enhancement of financial knowledge.

This condition implies that modern financial technology enables individuals to directly utilize digital financial services without necessarily possessing extensive financial knowledge. The ease of use of financial applications, digital payment systems, and automated transaction services allows technology access to directly improve individuals' financial capabilities.

The total effect analysis further confirms the dominant influence of technology access on financial literacy. Under the regime ($X_1 \leq K_1$), the total effect reaches 0.747, while under the regime ($X_1 > K_1$), the total effect decreases to 0.424. These results further emphasize the nonlinear relationship pattern and demonstrate that the effectiveness of technology access differs across levels of access.

The empirical findings demonstrate that the semiparametric path analysis model combined with double jackknife inference provides flexible and stable estimation while being more capable of capturing nonlinear relationships compared to conventional linear approaches. The integration of truncated spline modeling and double resampling inference also offers a more robust analytical framework for modeling complex causal relationships.

3.7 Discussion

The results of this study provide methodological and empirical contributions to the development of semiparametric path analysis for modeling the complex relationships among access to technology, financial knowledge, and financial literacy. From a methodological perspective, simulation results show that the double jackknife resampling approach yields more stable estimators than the single resampling method, particularly for medium sample sizes and nonlinear model structures. Lower RMSE and bias values, along with a higher coverage probability, indicate that adding a resampling layer improves estimation precision and reduces inference variability. These findings align with previous research stating that repeated resampling procedures can enhance estimator robustness in semiparametric models when classical distribution assumptions are difficult to satisfy (Rizqia et al., 2026).

The relatively stable estimation bias across various sample sizes and error variances also indicates that the spline approach used is sufficiently robust despite approximation errors resulting from knot selection. These findings are consistent with previous semiparametric spline studies stating that bias generally stems from a trade-off between model flexibility and the smoothness level of the function (Rohma et al., 2025). However, the magnitude of the bias in this study remains relatively small, suggesting that the spline specification used is capable of representing the underlying nonlinear patterns in the relationships between variables.

Empirically, the results of the Ramsey RESET test confirm that the relationship between technology access and financial literacy cannot be optimally explained using a pure linear model. This indicates that the formation of financial literacy is heterogeneous and influenced by changing relationship patterns at specific levels of technology access. Therefore, the use of truncated splines is relevant because they can capture changes in the structure of relationships that cannot be identified by conventional linear models.

Estimation results indicate that technology access has a significant positive effect on financial knowledge and financial literacy. This finding supports the research by Yang et al. (2023), which shows that access to digital technology expands access to financial information, enhances understanding of financial products, and accelerates the financial decision-making process. However, the effect of financial knowledge on financial literacy in this study is relatively small, indicating that improvements in financial literacy are not solely determined by cognitive aspects but are also influenced by the ease of using digital financial services and direct experience with financial technology (Kurniasari et al., 2025).

The main finding of this study is the presence of a nonlinear relationship between technology access and financial literacy, marked by a knot at $X_1 = 3.667$. This pattern indicates a diminishing marginal effect, where increased technology access provides greater benefits to groups with low access levels, but the marginal effect diminishes after passing a certain point. This finding expands upon previous studies that generally assumed a linear relationship between technology and financial literacy. Substantively, these results suggest that the benefits of technology do not increase proportionally without limit, as at higher levels of access, individuals tend to have already reached a relatively optimal state of information (Kulshrestha, 2023).

Interestingly, the indirect effect of technology access on financial literacy through financial knowledge was found to be insignificant. This suggests that technology influences financial literacy more through direct mechanisms than through increased formal financial knowledge. These findings reinforce the arguments of Lusardi and Mitchell (2014) and Morris et al. (2022) that the development of increasingly simple and intuitive digital financial services enables individuals to make financial decisions without requiring a high level of financial knowledge. In other words, the digitization of financial services can reduce cognitive barriers to the use of financial products.

From a practical perspective, the results of this study suggest that policies to improve financial literacy should prioritize expanding access to technology, particularly among groups with low access levels, as these groups derive greater marginal benefits. Additionally, the integration of truncated splines and double resampling jackknife inference indicates that the proposed approach provides a more flexible and robust methodological framework for modeling complex structural relationships compared to linear approaches.

4. CONCLUSION

This study highlights the importance of semiparametric approaches in path analysis when the assumption of linearity is not fully met. By integrating truncated splines and double jackknife resampling, the developed model is capable of simultaneously accommodating both linear and nonlinear relationships and produces more robust inferences compared to parametric approaches. Simulation results indicate that the double jackknife method provides the best estimation performance, particularly for medium sample sizes with a single nonlinear relationship, as evidenced by lower RMSE and bias, as well as a higher coverage probability. Empirical results show that access to technology has a positive effect on financial knowledge and financial literacy, while the indirect effect through financial knowledge is not significant. Additionally, the truncated spline successfully identified a diminishing marginal effect, where the impact of technology access on financial literacy decreases after passing a certain knot point. This finding confirms that relationships among variables in financial literacy are not always linear and require a more flexible approach to capture complex relationship patterns. This study still has limitations regarding the number of nonlinear relationships modeled and the sample size used in the simulation. Future research could develop models with multiple nonlinear components, a larger sample size, and compare nonparametric approaches with other resampling strategies to evaluate model robustness more comprehensively.

5. ACKNOWLEDGMENT

The author expresses sincere gratitude to all respondents who generously participated in this study and contributed valuable data and insights essential to the completion of this research. Their willingness to provide information has significantly supported the empirical foundation and overall quality of this work.

REFERENCES

Amanda, D., A. Iriany, A. A. R. Fernandes, and Solimun (2024). Analysis of Path Nonparametric Truncated Spline Maximum Cubic Order in Banking Credit of Risk Behavior Model. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, **18**(4); 2639–2652

Asare, M. B., A. Boateng, E. T. Mensah, and D. Maposa (2024). Parameter Estimation of Ornstein-Uhlenbeck Process Using Resampling Estimation Techniques: A Simulation Study. In P. e. a. Vasant, editor, *Intelligent Computing and Optimization*, volume 1169 of *Lecture Notes in Networks and Systems*. Springer

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

Christodoulou-Volos, C. and D. Tserkezos (2023). Sensitivity of the Ramsey's Regression Specification Error Term Test on the Degree of Nonlinearity of the Functional Form. *Journal of Applied Economic Sciences*, **18**(1); 5–10

Efendi, E. C. L., A. A. R. Fernandes, and M. B. T. Mitakda (2021). Modeling of Path Nonparametric Truncated Spline Linear, Quadratic, and Cubic in Model on Time Paying Bank Credit. *WSEAS Transactions on Business and Economics*, **18**; 51–60

Fernandes, A. A. R. and Solimun (2021). *Analisis Regresi dalam Pendekatan Fleksibel*. UB Press

Fernandes, A. A. R., Solimun, L. Muflikhah, A. Alifa, E. Krisnawati, N. M. A. A. Badung, and E. C. L. Efendi (2022). Nonparametric Path Analysis on Consumer Satisfaction and Consumer Engagement in PT Pertamina. *WSEAS Transactions on Mathematics*, **21**; 17–22

Grapentine, T. (2000). Path Analysis vs Structural Equation Modeling. *Marketing Research*, **12**(3); 13–20

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson (2019). *Multivariate Data Analysis*. Cengage Learning, 8th edition

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition

Hidayat, M. F., A. A. R. Fernandes, and Solimun (2019). Estimation of Truncated Spline Function in Non-Parametric Path Analysis Based on Weighted Least Square (WLS). *IOP Conference Series: Materials Science and Engineering*, **546**; 052027

James, G., D. Witten, T. Hastie, and R. Tibshirani (2023). *An Introduction to Statistical Learning with Applications in R*. Springer

Junianto, F. H., A. A. R. Fernandes, Solimun, and A. B. Astuti (2025). Monte Carlo-Based Bayesian Path Analysis for Modeling Indirect Financial Effects on Literacy in Emerging Markets. *Journal of Economic Integration*, **40**(4); 677–696

Kulshrestha, S. (2023). The Role of Financial Technology in Enhancing Financial Literacy and Inclusion Among Low-Income Households in India. *International Journal of Research in Marketing Management and Sales*, **5**(1); 25–30

Kurniasari, F., N. Abd Hamid, and E. D. Lestari (2025). Unraveling the Impact of Financial Literacy, Financial Technology Adoption, and Access to Finance on Small Medium Enterprises Business Performance and Sustainability: A Serial Mediation Model. *Cogent Business & Management*, **12**(1)

Lusardi, A. and O. S. Mitchell (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, **52**(1); 5–44

Maatouk, H., D. Rullière, and X. Bay (2024). Truncated Multivariate Normal Distribution Under Nonlinear Constraints. HAL Open Science preprint

MacKinnon, J. G. (2006). Bootstrap Methods in Econometrics. *The Economic Record*, **82**(s1); S2–S18

Morris, T., S. Maillet, and V. Koffi (2022). Financial Knowledge, Financial Confidence and Learning Capacity on Financial Behavior: A Canadian Study. *Cogent Social Sciences*, **8**(1); 1996919

Papalia, M. F., Solimun, and N. Nurjannah (2023). Comparison of Resampling Efficiency Levels of Jackknife and Double Jackknife in Path Analysis. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, **17**(2); 807–818

Ramsay, J. O. (1988). Monotone Regression Splines in Action. *Statistical Science*, **3**(4); 425–441

Rizqia, A., Solimun, N. Nurjannah, K. Hidayat, and F. Junianto (2026). Quadratic and Truncated Spline Structural Equation Modeling With Double Bootstrap in the Waste Management Economy. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*

Robert, C. P. and G. Casella (2010). *Monte Carlo Statistical Methods*. Springer, 2 edition

Rohma, U., A. A. R. Fernandes, S. Astutik, and Solimun (2025). Development of Nonparametric Path Function Using Hybrid Truncated Spline and Kernel for Modeling Waste-to-Economic Value Behavior. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, **19**(1); 331–344

Solimun, A. A. R. Fernandes, and I. Rahmawati (2021). Research on Structural Flexibility and Acceptance Model (SFAM) Reconstruction Based on Disruption Innovation in the Social Humanities and Education Sector. *WSEAS Transactions on Mathematics*, **20**; 657–675

Suriaslan, A. S., I. N. Budiantara, and V. Ratnasari (2025). Nonparametric Regression Estimation Using Multivariable Truncated Splines for Binary Response Data. *MethodsX*, **14**

Vinod, H. D. (2022). Risk and Financial Management Kernel Regression Coefficients for Practical Significance

Wood, S. N. (2026). Generalized Additive Models. *Annual Review of Statistics and Its Application*, **29**; 10

Yang, J., Y. Wu, and B. Huang (2023). Digital Finance and Financial Literacy: Evidence from Chinese Households. *Journal of Banking & Finance*, **156**; 107005

