CAUCHY –Jurnal Matematika Murni dan Aplikasi
Volume 10(1) (2025), Pages 250-262
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Development of Semiparametric Truncated Spline Logistic Path Analysis

Sasi Wilujeng Sri Rejeki[1], Solimun, Nurjannah[1], Muhammad Ohid Ullah[2*]

[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Indonesia
[2]Department of Statistics, Shahjalal University of Science and Technology Sylhet, Bangladesh

Email: sasiwsr@student.ub.ac.id; ohid-sta@sust.edu*

ABSTRACT

Logistic path analysis extends logistic regression by incorporating intervening variables, addressing the limitations of linearity assumptions through nonparametric models like spline regression. However, this study develops a semiparametric truncated spline logistic path analysis to accommodate linear and nonlinear relationships, considering direct and indirect effects of intervening variables. The model is applied to analyze the impact of price volatility and human resource quality on farmer welfare, with farmer productivity as an intervening variable. It assumes a nonlinear relationship between price volatility and productivity/welfare, while other relationships are linear. This development was applied to secondary data collected through questionnaires from farmer group members in Bali Province, which were analyzed using a semiparametric truncated spline logistic path model. Optimal knots were determined using the lowest GCV value. The results show that the model effectively captures changes in data patterns, providing robust parameter estimates. Hypothesis testing highlights significant differences in the effectiveness of linear and nonlinear relationships. The use of truncated splines offers critical insights into variable interactions and enhances model reliability, making it a valuable tool for analyzing complex agricultural systems and informing policies to improve farmer welfare and productivity.

Keywords: Farmer Inclusiveness; Human Resource Development in Agriculture; Agricultural Price Volatility; Semiparametric Logistic Path Analysis; Truncated Spline Analysis

INTRODUCTION

Logistic path analysis is a method for analyzing the relationships between exogenous variables and pure endogenous variables through intervening variables, where both the pure endogenous variables and intervening variables are categorical [1]. Logistic path analysis is an extension of logistic regression analysis that incorporates intervening variables. Logistic regression models are a method used to explain the relationship between categorical response variables and predictor variables, which can be either continuous or categorical [2]. Logistic regression models facilitate the analysis of binary

Submitted: November 23, 2024 Reviewed: December 05, 2024 Accepted: January 27, 2025
DOI: http://dx.doi.org/10.18860/ca.v10i1.29979

dependent variables with one or more independent variables, but they rely on the assumption of a linear relationship on the logit scale, which may not always capture more complex data patterns [3]. Advances in regression analysis methods aim to address the limitations of the linearity assumption in logistic regression models by introducing nonparametric models, such as nonparametric logistic regression with a spline approach. The relationship between independent and dependent variables to vary according to data patterns without enforcing a linear relationship [4]. Splines enable models to accommodate nonlinear changes, providing more accurate estimates in situations where the relationship between variables is not predefined. For example, the study by [5] demonstrated that a binary logistic regression model with a single knot spline achieved high accuracy in classifying nutritional status. Similarly, [6]showed that ordinal logistic regression models with truncated spline estimators effectively captured changes in data behavior across different intervals.

Further research by [7] highlighted the flexibility of truncated spline estimators in handling complex data through the selection of optimal knots. This knot selection, conducted using the Generalized Cross Validation (GCV) method, significantly improved model performance. The approach is beneficial in addressing patterns of variable relationships that cannot be captured by conventional binary logistic regression models, as explained by [8]. Although nonparametric logistic regression with truncated splines excels at identifying nonlinear patterns, it is not yet sufficient for analyzing causal relationships involving multiple pathways. This limitation underscores the need for further development of path analysis [9].

Based on previous findings, this study aims to develop a semiparametric truncated spline logistic path analysis as a flexible approach to modeling complex and nonlinear relationships, particularly in cases involving intervening variables. This method can be applied to issues such as farmer inclusivity in Bali, which aims to integrate small-scale farmers into agricultural value chains [10]. Bali is intriguing to study because it has a unique agricultural culture called Subak, a traditional institution that plays a critical role in managing water resources for agriculture but faces challenges that hinder the growth of the agricultural sector. [11] emphasized the importance of community-based policies and stakeholder collaboration to enhance farmer inclusivity while respecting Bali's cultural landscape. Thus, Balinese farmers should be included in agricultural programs and policies to promote their inclusivity. However, agricultural management policies need to consider the characteristics of the farmers, particularly in terms of their welfare.

The Coordinating Ministry for Economic Affairs of the Republic of Indonesia on 2021 stated that the welfare of farmers integrated within inclusivity initiatives has a direct impact on national food security, as prosperous farmers are better positioned to enhance the quality and productivity of their agricultural activities [12]. High productivity provides opportunities for farmers to increase their income and overall welfare. The study by [13] states that increased productivity can enhance farmers' income.Thus, policies that support linear improvements in productivity are essential for enhancing farmer welfare. Additionally, [14] demonstrated that food price volatility affects productivity, while [15] emphasized the positive impact of human resource quality on agricultural productivity. However, if there are indications of extreme price volatility, it is possible that the relationship between price volatility and farmer productivity is nonlinear. Meanwhile, in some cases, the quality of farmers shows a linear relationship with productivity and farmer welfare. Consequently, semiparametric truncated spline logistic path analysis is expected to capture the relationships between these variables, both linear and nonlinear, by considering intervening variables and facilitating more effective policymaking for farmers in Bali. The use of semiparametric truncated spline logistic analysis has not been found in previous studies, making it a novelty in this research.

METHODS

The data used in this study consists of secondary data. The secondary data was obtained from a research grant conducted by Solimun et al. [16]. This data was collected through a questionnaire survey targeting farmers who are members of farmer groups in Bali Province, an operational area of a state-owned fertilizer company. The population includes 477,439 individuals distributed across 12,323 farmer groups. These farmers are group members, landowners, and have at least a senior high school (or equivalent) level of education. The sample size was determined using the Slovin formula as follows.

$$n = \frac{N}{(1+N(e)^2)} \tag{1}$$

wich

n : sample size

N : population size

e : tolerable error rate (5-10%)

The population (N) in this study is as many as 12,323 farmer groups from Bali Province who use fertilizer from one of the SOEs. With an error rate of 7.5%, the sample can be calculated as follows.

$$n = \frac{12.323}{(1+12.323(0,075)^2)} = 175,25 \approx 176 \text{ Farmer Group}$$

Based on these calculations, the sample used for this study was rounded up into 176 farmer groups. Each farmer group is taken as many as one member. Thus, many samples n = 176 farmers.

The variables used in this study are Price Volatility (X_1) and Quality of Human Resources (X_2) as exogenous variables in the form of farmers' perceptions assessed with likert-scale indicators. The Farmer Income (Y_1) and Farmer Welfare Status (Y_2) variables are endogenous variables on a binary ordinal scale with categories of 0 (low) and 1 (high). The conceptual model of the research is presented in **Figure 1**.

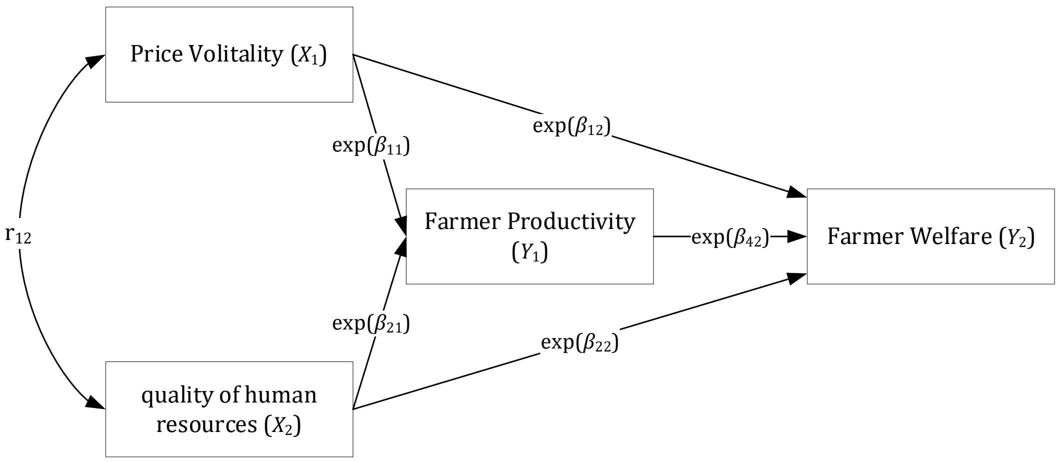

Figure 1. Research Conceptual Model

Based on the conceptual model above, a Semiparametric Truncated Spline Logistic

Path Analysis with an indication of a nonlinear relationship between X_1 to Y_1 and Y_2 will be developed as shown in Equation (2).

$$\pi_{1i}\ (X_{1i},(X_{1i}-K_{11})_+,X_{2i}) = \frac{\exp(\beta_{01}+\beta_{11}\ X_{1i}+\beta_{21}(X_{1i}-K_{11})_+,+\beta_{31}X_{2i})}{1+\exp(\beta_{01}+\beta_{11}\ X_{1i}+\beta_{21}(X_{1i}-K_{11})_+,+\beta_{31}X_{2i})}$$

$$\pi_{2i}\ (X_{1i},(X_{1i}-K_{12})_+,X_{2i},Y_{1i}) = \frac{\exp(\beta_{02}+\beta_{12}\ X_{1i}+\beta_{22}(X_{1i}-K_{11})_+,+\beta_{32}X_{2i}+\beta_{42}Y_{1i})}{1+\exp(\beta_{02}+\beta_{12}\ X_{1i}+\beta_{22}(X_{1i}-K_{11})_+,+\beta_{32}X_{2i}+\beta_{42}Y_{1i})} \quad (2)$$

To facilitate the calculation of parameter estimation, it can be written in the logit model in the equation (3).

$$\log\left(\frac{\pi_{1i}}{1-\pi_{1i}}\right)=\beta_{01}+\beta_{11}\ X_{1i}+\beta_{21}(X_{1i}-K_{11})_+,+\beta_{31}X_{2i}$$

$$\log\left(\frac{\pi_{2i}}{1-\pi_{2i}}\right)=\beta_{02}+\beta_{12}\ X_{1i}+\beta_{22}(X_{1i}-K_{12})_+,+\beta_{32}X_{2i}+\beta_{42}Y_{1i} \quad (3)$$

with truncated function:

$$(X_{1i}-K_{11})_+=\begin{cases}(X_{1i}-K_{11}) & ;X_{1i}\ge K_{11}\\ 0 & ;X_{1i}<K_{11}\end{cases}$$

$$(X_{1i}-K_{12})_+=\begin{cases}(X_{1i}-K_{12}) & ;X_{1i}\ge K_{12}\\ 0 & ;X_{1i}<K_{12}\end{cases}$$

The selection of the best knots leads more to parsimony or simplicity of the model. The optimal knot selection is carried out based on the value of the smallest GCV (Generalized Cross Validation) [17]. GCV for a spline model with **K** number of knots is defined as the ratio between the Mean Squared Error (MSE) for a model with **K** knots to the square of the number of diagonal (trace) of **the I-A(K)** matrix, normalized by the sample size *n*. Where **A(K)** is the hat matrix for the model with **K** knots, **I** is the identity matrix, and *n* is the number of observations in the dataset. The GCV formula is as shown in equation (4).

$$GCV(\mathbf{K})=\frac{MSE(\mathbf{K})}{\left[n^{-1}trace(\mathbf{I}-\mathbf{A}(\mathbf{K}))\right]^2} \quad (4)$$

Logistics line parameters can be estimated by the Maximum Likelihood Estimation (MLE) method [1]. The MLE method is a technique used to estimate the parameters of a probability distribution based on the observed data. MLE aims to find parameter values that maximize the likelihood function, i.e. a function that states the likelihood that the observed data appears based on certain parameters of the probability model [18]. If θ is the estimated parameter vector and X is the observation data, then the likelihood function $L(\theta|X)$ represents the probability of X data given the parameter θ. Thus, likelihood θ from the observation data can be expressed in equation (5).

$$\mathrm{L}(\theta\,|\,X)\propto \mathrm{P}(X,\theta) \quad (5)$$

The purpose of MLE is to find $\hat{\theta}$ which maximizes $\mathrm{L}(\theta\,|\,\mathrm{X})$. In practice, it is easier to use the concept of working with log-likelihood functions, which is the logarithm of the likelihood function as in equation (6).

$$l(\theta\,|\,X)=\mathrm{In}\ \mathrm{L}(\theta\,|\,X) \quad (6)$$

To achieve value $\hat{\theta}$ which maximizes $l(\theta\,|\,\mathrm{X})$ by finding the first derivative of log-likelihood to and equalizing it to zero as in equation (7).

$$\frac{\partial l(\theta\,|\,X)}{\partial\theta}=0 \quad (7)$$

To ensure the value of $\hat{\theta}$ has maximized $l(\theta \mid X)$, can check the second derivative of log-likelihood with the value $\hat{\theta}$ negative values [19]. However, in some models it is not possible to estimate the parameters directly, so an iterative method is needed to gradually improve the estimate to get closer to the value $\hat{\theta}$ the most convergent. Convergence is achieved when the change in parameter estimation between iterations is very small. The iterative method that is often used is the Newton-Raphson method which refers to the first and second derivatives of the log-likelihood function with iterations

$$\theta^{t+1} = \theta^{t} - (H^{t})^{-1} q^{t}$$

wich θ is a parameter vector where t indicates the t-th iteration, q^t is the gradient vector of the log-likelihood function at the t-th iteration with equation (8).

$$q^{t} = \frac{\partial l(\theta \mid X)}{\partial \theta} \tag{8}$$

Wich $l(\theta^t)$ is the log-likelihood of the parameter θ on the iteration t. While, H^t is the Hessian Matrix as the second derivative of the t-th iteration with the element of equation (9).

$$H^{t} = \frac{\partial^{2} l(\theta^{t})}{\partial \theta \, \partial \theta^{t}} \tag{9}$$

When a response variable on logistic path analysis with paramater β is a binary variable or a dichotomy, meaning that both response variables consist of two categories, namely "success" $(Y_{ji} = 1) \rightarrow P(Y_{ji} = 1) = \pi_{ji}$ or "failed" $(Y_{ji} = 0) \rightarrow P(Y_{ji} = 0) = 1 - \pi_{ji}$ dimana $j = 1,2$ dan $i = 1,2,3,\ldots,n$ merupakan indeks yang menunjukkan, then the variable Y_{ji} follows the Bernoulli distribution. It is stated that binary variables are independent of each other, so the sum of the binary variables will have a binomial spread. It is known that the model has a probability density function in the logistic path model as shown in Equation (10).

$$\begin{aligned} &\prod_{i=1}^{n} \left(\pi_{1i}^{Y_{1i}} (1-\pi_{1i})^{1-Y_{1i}} \right) \left(\pi_{2i}^{Y_{2i}} (1-\pi_{2i})^{1-Y_{2i}} \right) \\ &= \left\{ \prod_{i=1}^{n} (1-\pi_{1i})(1-\pi_{2i}) \right\} \left\{ \exp\left(\sum_{i=1}^{n} Y_{1i} \log_e \left(\frac{\pi_{1i}}{1-\pi_{1i}} \right) + Y_{2i} \log_e \left(\frac{\pi_{2i}}{1-\pi_{2i}} \right) \right) \right\} \end{aligned} \tag{10}$$

The odds ratio provides a valid measure of effect so that it is able to describe the ratio of the chance of success or not of the response variable affected by the predictor variable [1]. The odds of an event are defined as the probability of a successful outcome divided by the probability of a failure event as calculated by the formula in equation (11).

$$\text{Odds}(\beta) = \frac{\pi_{\text{i}}}{(1-\pi_{\text{i}})} = \exp(\beta) \tag{11}$$

wich

π_i : probability of success $(Y = 1)$

$1 - \pi_i$: probability of failure $(Y = 0)$

The goodness of the model in the analysis of semiparametric logistic paths can be obtained in a gradual way, namely deviance to test the feasibility of a general linear model that shows a higher number size indicating a worse fit. The Total Determination Coefficient is the total diversity of data that can be explained by the model can be shown by the total determination coefficient using equations (12)[20].

$$R_{total}^{2} = 1 - P_{e1}^{2} P_{e2}^{2} \ldots P_{e_M}^{2} \tag{12}$$

The calculation of the residual effect (*error*) can be done using equation (13).

$$P^2_{em} = 1 - R^2_{MF_m} \quad (13)$$

The calculation of the determination coefficient can be done using equation (14).

$$R^2_{MF_m} = 1 - \frac{l_m}{l_{0m}} \quad (14)$$

wich:

R^2_{total}	:	Total determination coefficient
$P^2_{e_M}$	:	Quadratic influence of the remainder of the *i*-th model (*i* = 1, 2, ..., *m*)
$R^2_{MF_m}$	:	The determination coefficient of the *i*-th model (*i* = 1, 2, ..., *m*)
l_m	:	Loglikelihood full of the *i*-th model (*i* = 1, 2, ..., *m*)
l_{0m}	:	Loglikelihood null of the *i*-th model (*i* = 1, 2, ..., *m*)

The total determination coefficient has a value ranging from 0% to 100%. The greater the value of the total determination coefficient close to 100%, the better the model is said to be.

RESULTS AND DISCUSSION

Development of Estimation of Functional Parameters of Semiparametric Truncated Spline Logistics Line

The truncated spline semiparametric logistics path model is a development of the classical logistics path model that is used when there is a relationship between exogenous variables and the probability of occurrence of events in endogenous variables that are not linear. The linearity of the relationship between exogenous variables and the logit (log-odds) values of endogenous variables can be observed through the scatterplot as follows.

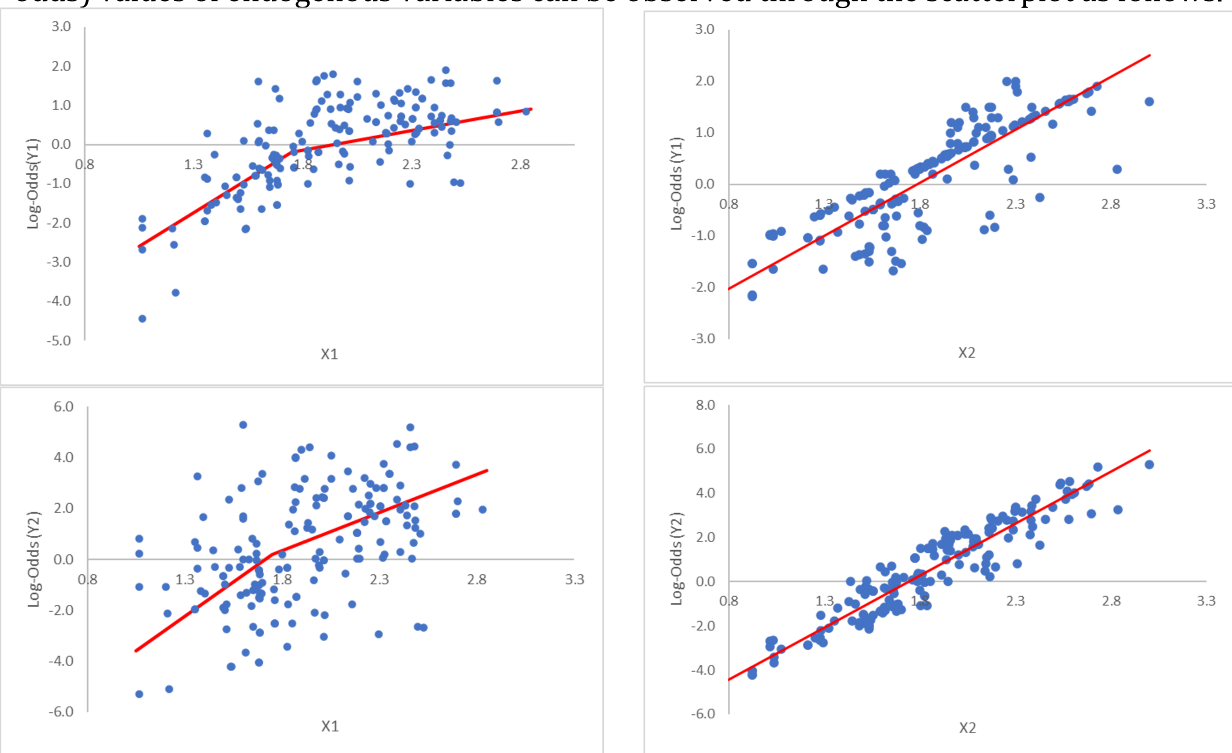

Figure 2. Scatterplots illustrate the relationships between Price Volatility (X1) and Quality of Human

Resources (X2) with the log-odds of Farmer Productivity (Y1) and Farmer Welfare (Y2)

Based on the scatterplot in Figure 2, it shows a relationship between the Price Volatility variable (X_1) and the log-odds of the Farmer Productivity variable (Y_1), as well as between the Price Volatility variable (X_1) and the log-odds of the Farmer Welfare variable (Y_2). Thus, the semiparametric truncated spline logistic path analysis is appropriate to be applied to this data. then the path equation can be written as equation (2) and equation (3). Equation (3) can be written in the form of a matrix as equation (15).

$$\begin{bmatrix} \log_e\left(\frac{\pi_{11}}{1-\pi_{11}}\right) \\ \log_e\left(\frac{\pi_{12}}{1-\pi_{12}}\right) \\ \vdots \\ \log_e\left(\frac{\pi_{1n}}{1-\pi_{1n}}\right) \\ \log_e\left(\frac{\pi_{21}}{1-\pi_{21}}\right) \\ \log_e\left(\frac{\pi_{21}}{1-\pi_{21}}\right) \\ \vdots \\ \log_e\left(\frac{\pi_{2i}}{1-\pi_{2i}}\right) \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & (X_{11}-K_{11})_+ & X_{21} & 0 & 0 & 0 & 0 & 0 \\ 1 & X_{12} & (X_{12}-K_{11})_+ & X_{22} & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1i} & (X_{1i}-K_{11})_+ & X_{2i} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & X_{11} & (X_{11}-K_{12})_+ & X_{21} & Y_{11} \\ 0 & 0 & 0 & 0 & 1 & X_{12} & (X_{12}-K_{12})_+ & X_{22} & Y_{12} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & X_{1i} & (X_{13}-K_{12})_+ & X_{2i} & Y_{1i} \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{02} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \end{bmatrix}$$

atau dapat dituliskan sebagai notasi matriks berikut.

$$\boldsymbol{logit}_{n\times 1} = \boldsymbol{X}'_{n\times p}\boldsymbol{\beta}_{p\times 1} \qquad (15)$$

with ***logit*** being the matrix of log-odds ratios for the endogenous variable. Estimating the parameters of the truncated spline semiparametric logistics path function using the MLE method, with *the likelihood* function as shown in equation (16).

$$\prod_{i=1}^{n}\left(\pi_{1i}^{Y_{1i}}(1-\pi_{1i})^{1-Y_{1i}}\right)\left(\pi_{2i}^{Y_{2i}}(1-\pi_{2i})^{1-Y_{2i}}\right)$$
$$=\left\{\prod_{i=1}^{n}(1-\pi_{1i})(1-\pi_{2i})\right\}\left\{\exp\left(\sum_{i=1}^{n}Y_{1i}\log_e\left(\frac{\pi_{1i}}{1-\pi_{1i}}\right)+Y_{2i}\log_e\left(\frac{\pi_{2i}}{1-\pi_{2i}}\right)\right)\right\} \qquad (16)$$

Which equation (3) can be substituted into the following equation (16) to be equation (17).

$$\prod_{i=1}^{n}\left(\pi_{1i}^{Y_{1i}}(1-\pi_{1i})^{1-Y_{1i}}\right)\left(\pi_{2i}^{Y_{2i}}(1-\pi_{2i})^{1-Y_{2i}}\right)$$
$$=\left\{\prod_{i=1}^{n}(1-\pi_{1i})(1-\pi_{2i})\right\}$$
$$\left\{\exp\left(\sum_{i=1}^{n}Y_{1i}\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}+K_{11})_+ +\beta_{31}X_{2i}\right)+Y_{2i}\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}+K_{12})_+ +\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)\right)\right\}$$

while

$$(1-\pi_{1i}) = \left(1+\exp\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}-K_{11})_{+}+\beta_{31}X_{2i}\right)\right)^{-1}$$

$$(1-\pi_{2i}) = \left(1+\exp\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}-K_{12})_{+}+\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)\right)^{-1} \quad (17)$$

thus the following equation is obtained

$$\prod_{i=1}^{l}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}$$

$$=\left\{\prod_{i=1}^{l}\left(1+\exp\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}+K_{11})_{+}+\beta_{31}X_{2i}\right)^{-1}\right)\left(1+\exp\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}+K_{12})_{+}+\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)^{-1}\right)\right\}$$

$$\left\{\exp\left(\sum_{i=1}^{n}Y_{1i}\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}+K_{11})_{+}+\beta_{31}X_{2i}\right)+Y_{2i}\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}+K_{12})_{+}+\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)\right)\right\}$$

$$=\left\{\prod_{i=1}^{l}\left(1+\exp\left(\text{logit}_{Y_{1i}}\right)\right)^{-1}\left(1+\exp\left(\text{logit}_{Y_{2i}}\right)\right)^{-1}\right\}$$

$$\left\{\exp\left(\sum_{i=1}^{n}Y_{1i}\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}+K_{11})_{+}+\beta_{31}X_{2i}\right)+Y_{2i}\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}+K_{12})_{+}+\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)\right)\right\}$$

$$=\left\{\prod_{i=1}^{l}\left(1+\left(\exp\left(\text{logit}_{Y_{1i}}\right)^{-1}\right)\right)\left(1+\left(\exp\left(\text{logit}_{Y_{2i}}\right)^{-1}\right)\right)\right\}$$

$$\left\{\exp\left(\sum_{i=1}^{n}Y_{1i}\left(\beta_{01}+\beta_{11}X_{1i}+\beta_{21}(X_{1i}+K_{11})_{+}+\beta_{31}X_{2i}\right)+Y_{2i}\left(\beta_{02}+\beta_{12}X_{1i}+\beta_{22}(X_{1i}+K_{12})_{+}+\beta_{32}X_{2i}+\beta_{42}Y_{1i}\right)\right)\right\}$$

or can be written in matrix notation as follows.

$$\prod_{i=1}^{l}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}$$

$$=\left\{1+\exp\left[\boldsymbol{logit}^{-1}\right]\right\}\left\{\exp\left[\boldsymbol{Y}_{\boldsymbol{n}\times\boldsymbol{n}}\boldsymbol{X}'_{\boldsymbol{n}\times\boldsymbol{9}}\boldsymbol{\beta}_{\boldsymbol{9}\times\boldsymbol{1}}\right]_{n\times\boldsymbol{1}}\right\} \quad (18)$$

$$=\left\{1+\exp\left[\left[\boldsymbol{X}'_{\boldsymbol{n}\times\boldsymbol{9}}\boldsymbol{\beta}_{\boldsymbol{9}\times\boldsymbol{1}}\right]^{-1}\right]\right\}\left\{\exp\left[\boldsymbol{Y}_{\boldsymbol{n}\times\boldsymbol{n}}\boldsymbol{X}'_{\boldsymbol{n}\times\boldsymbol{9}}\boldsymbol{\beta}_{\boldsymbol{9}\times\boldsymbol{1}}\right]_{n\times\boldsymbol{1}}\right\}$$

which

$$\boldsymbol{Y}_{\boldsymbol{n}\times\boldsymbol{n}}=\begin{bmatrix} Y_{11} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & Y_{12} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & Y_{1n} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & Y_{21} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & Y_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & Y_{2n} \end{bmatrix}$$

To estimate the parameters of the truncated spline nonparametric logistic regression, the next process is to deduce the log-likelihood function to the parameter and equal to zero as in equation (19).

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\frac{\partial\left\{1+\boldsymbol{exp}\left[\left[\boldsymbol{X}'_{\boldsymbol{n}\times\boldsymbol{9}}\boldsymbol{\beta}_{\boldsymbol{9}\times\boldsymbol{1}}\right]^{\boldsymbol{-1}}\right]\right\}\left\{\boldsymbol{exp}\left[\boldsymbol{Y}_{\boldsymbol{n}\times\boldsymbol{n}}\boldsymbol{X}'_{\boldsymbol{n}\times\boldsymbol{9}}\boldsymbol{\beta}_{\boldsymbol{9}\times\boldsymbol{1}}\right]_{\boldsymbol{n}\times\boldsymbol{1}}\right\}}{\partial\boldsymbol{\beta}}=0 \quad (19)$$

To get the beta estimator value, the feeding is done Newton *Raphson*'s algorithm with iterations $\boldsymbol{\beta}^{t+1}=\boldsymbol{\beta}^{t}-(\boldsymbol{H}^{t})^{-1}\boldsymbol{q}^{t}$, shows the model parameter vector, t shows the t-th iteration, $\boldsymbol{q^t}$ is a vector that is the first derivative of the likelihood function and $\boldsymbol{H^t}$ is a Hessian matrix which is the second derivative of the likelihood function.

Results of Estimation of the Function of Truncated Spline Semiparametric Logistics Lines on Farmer Inclusivity Data

The semiparametric logistics path model can be described as shown in equation (2). The estimation of parameters in the truncated spline semiparametric logistics path function is carried out on the logit model listed in equation (3) with the MLE method that has been described in equations (16) to equations (19).

The selection of optimal knots can be seen through *the goodness of fit* in each relationship between exogenous variables and the chance of endogenous variables in **Table 1.**

Table 1. Results of Optimal Knot Selection

Nonparametric Relationship	**Knots**	***Goodness of Fit***				
		GCV	**Accuracy**	**Specificity**	**Sensitifity**	$\mathbf{R^2_{total}}$
$\boldsymbol{X_1 \rightarrow Y_1}$	K_{11} = 1,71	0,259	72,72%	73,60%	71,61%	72,86%
$\boldsymbol{X_2 \rightarrow Y_1}$	K_{21} = 2,52	0,250	73,30%	74,11%	72,26%	73,59%
$\boldsymbol{X_1 \rightarrow Y_2}$	K_{12} = 1,71	0,259	72,73%	73,61%	71,62%	72,86%
$\boldsymbol{X_2 \rightarrow Y_2}$	K_{22} = 2,76	0,251	56,82%	69,47%	52,14%	73,80%
$\boldsymbol{X_1 \rightarrow Y_1}$ $\boldsymbol{X_1 \rightarrow Y_2}$	K_{11} = 1,71 K_{12} = 1,71	0,211	72,44%	71,15%	73,47%	72,29%
$\boldsymbol{X_1 \rightarrow Y_1}$ $\boldsymbol{X_2 \rightarrow Y_2}$	K_{11} = 1,71 K_{22}= 2,76	0,250	72,00%	72,96%	71,56%	71,19%
$\boldsymbol{X_2 \rightarrow Y_1}$ $\boldsymbol{X_1 \rightarrow Y_2}$	K_{21} = 1,71 K_{12} = 2,69	0,229	73,86%	74,12%	73,87%	73,87%
$\boldsymbol{X_2 \rightarrow Y_1}$ $\boldsymbol{X_2 \rightarrow Y_2}$	K_{21} = 1,71 K_{22} = 2,76	0,215	73,86%	74,12%	73,87%	73,87%
$\boldsymbol{X_1 \rightarrow Y_2}$ $\boldsymbol{X_2 \rightarrow Y_2}$	K12 = 1,71 K22 = 2,76	0,229	73,86%	74,12%	73,87%	73,87%

Based on **Table 1**, it shows that the optimal knot is in a non-linear relationship between variable Price Volatility (X_1) to variable opportunity Farmer Productivity(Y_1) and variable opportunity of Farmer Welfare (Y_2). This is because the goodness of fit value is above 70% with the lowest GCV value of 0.211 which indicates that the existence of an optimal knot as a fault pattern of the relationship between these variables can accurately classify endogenous variables. The following is the result of estimating the functional parameters of the truncated spline semiparametric logistics path with the optimal knots K_{11} = 1,71 and K_{12} = 1,71 in equation (20).

$$\log_e\left(\frac{\pi_{1i}}{1-\pi_{1i}}\right) = \text{-}10{,}110 + 4{,}354\,X_{1i} - 4{,}286(X_{1i}\text{-}1{,}71)_+ + 1{,}684X_{2i}$$
$$\log_e\left(\frac{\pi_{2i}}{1-\pi_{2i}}\right) = \text{-}8{,}908 + 0{,}851X_{1i} + 0{,}481(X_{1i}\text{-}1{,}71)_+ + 3{,}973X_{2i} + 1{,}093Y_{1i} \quad (20)$$

with truncated function:

$$\left(X_{1i}-1{,}71\right)_+ = \begin{cases} \left(X_{1i}-1{,}71\right) & ;\text{jika } X_{1i} \geq 1{,}71 \\ 0 & ;\text{jika } X_{1i} < 1{,}71 \end{cases}$$
$$\left(X_{2i}-1{,}71\right)_+ = \begin{cases} \left(X_{1i}-1{,}71\right) & ;\text{jika } X_{1i} \geq 1{,}71 \\ 0 & ;\text{jika } X_{1i} < 1{,}71 \end{cases}$$

Value Odds Ratio

The magnitude of the influence of the ordinal logistics path analysis can be seen through the odds ratio (OR) value obtained by the equation. The odds ratio values for each path are presented in equation (20).

Table 2. Odds Ratio Calculation Results

Relationship Between Variables			Coefficient	Odds Ratio	P-Value
Exogenous Variable	Endgenous Variable	Intervening Variable			
$\boldsymbol{X_1}$	Y_1	-	$\exp(\beta_{11})$ (X1 < 1,71)	77,789	<0,001**
$\boldsymbol{X_1}$	Y_1	-	$\exp(\beta_{11}+\beta_{21})$ ($X_1 \geq 1{,}71$)	1,070	0,006**
$\boldsymbol{X_1}$	Y_1	-	$\exp(\beta_{31})$	5,387	0,038*
$\boldsymbol{X_1}$	Y_2	-	$\exp(\beta_{12})$ ($X_1 < 1{,}71$)	2,342	0,579
$\boldsymbol{X_1}$	Y_2	-	$\exp(\beta_{12}+\beta_{22})$ ($X_1 \geq 1{,}71$)	3,789	0,825
$\boldsymbol{X_2}$	Y_2	-	$\exp(\beta_{32})$	53,143	<0,001**
$\boldsymbol{Y_1}$	Y_2	-	$\exp(\beta_{42})$	2,983	0,009**
$\boldsymbol{X_1}$	Y_2	Y_1	$\exp(\beta_{11}\times\beta_{42})$ ($X_1 < 1{,}71$)	116,630	<0,001**
$\boldsymbol{X_1}$	Y_2	Y_1	$\exp((\beta_{11}+\beta_{21})\times\beta_{42})$ ($X_1 \geq 1{,}71$)	1,077	<0,001**
$\boldsymbol{X_2}$	Y_2	Y_1	$\exp(\beta_{31}\times\beta_{42})$	6,303	<0,001**

Note: ** Significant at α = 0,05 and *Significant α = 0,1

Based on Table 2, the information obtained indicates that the odds ratio (OR) can be interpreted to show that the effect of the relationship between Price Volatility (X_1) and Farmer Productivity (Y_1) is significant, with an odds ratio of 77.789 under the knot condition. This odds ratio indicates that an increase in price volatility below the critical point is 77.789 times more likely to result in high farmer productivity compared to low productivity. This suggests that when fluctuations are not too extreme, farmers are more likely to provide optimal care for their crops, tailored to specific needs without facing significant challenges, thereby increasing their productivity.

In contrast, when Price Volatility (X_1) affects Farmer Productivity (Y_1) above the knot condition, the odds ratio is 1.070. This indicates that extreme price fluctuations are 1.070 times more likely to achieve high farmer productivity compared to low productivity. However, under non-extreme fluctuations, the productivity increase is higher than during extreme price fluctuations. This finding aligns with previous research by Martins-Filho & Toreto (2013), which stated that significant price fluctuations could lead to losses for farmers, while lower price fluctuations allow farmers to operate more stably and increase their agricultural yields.

Meanwhile, Price Volatility (X_1) does not significantly affect Farmer Welfare (Y_2) under either less or more than critical point conditions. This is because price volatility cannot directly impact farmers' welfare. However, the relationship between Price Volatility (X_1) and Farmer Welfare (Y_2) through Farmer Productivity ((Y_1) under the critical point condition shows a significant effect with an odds ratio of 116.630 for achieving high farmer welfare compared to low welfare. Similarly, under conditions less than the critical point, the odds ratio is 1.077, indicating that higher welfare is 1.077 times more likely when productivity is high than when it is low. This finding represents a novelty in the research, supported by the significant relationship between Farmer Productivity (Y_1) and Farmer Welfare (Y_2), with an odds ratio of 2.983.

The effect of Quality of Human Resources (X_2) on Farmer Productivity (Y_1) shows a significant impact, with an odds ratio of 5.387. This value indicates that an improvement in the Quality of Human Resources is 5.387 times more likely to achieve high productivity compared to low productivity. This finding is consistent with research by Oktavia et al. (2014), which stated that the quality of human resources significantly affects farmers' productivity. Improved skills and knowledge enable farmers to utilize resources more efficiently, contributing to increased productivity.

Moreover, the effect of Quality of Human Resources (X_2) on Farmer Welfare (Y_2) shows a significant relationship, with an odds ratio of 53.143. This value indicates that an improvement in the quality of farmers is 53.143 times more likely to achieve high farmer welfare compared to low welfare. The indirect relationship between Quality of Human Resources (X_2) and Farmer Welfare (Y_2) through Farmer Productivity (Y_1) also shows a significant effect, with an odds ratio of 6.303 times more likely to achieve high welfare compared to low welfare. The decline in the odds ratio for the indirect relationship compared to the direct effect is due to the possibility that as the quality of farmers improves, they seek additional jobs as alternative income sources to achieve better welfare. This is independent of agricultural productivity, which depends on external field conditions and results in income being realized more slowly by farmers.

CONCLUSIONS

Hypothesis testing on the function of the semiparametric logistic path analysis shows a significant difference in the effectiveness of the model between linear and non-linear relationships on certain variables. The use of truncated splines in hypothesis testing allows for the identification of knot where the relationships between variables change significantly, providing deeper insights into the influence of independent variables on dependent variables. Moreover, the insignificant direct relationship between the quality of human resources and farmer welfare can be addressed through a significant indirect relationship by considering farmer productivity as an intervening variable. These results are supported by the field conditions, which have indeed become a significant issue among Balinese farmers. This model shows quite good validity in classifying data with consistent accuracy across various scenarios, especially at smaller data proportions. Goodness of Fit shows that the truncated spline semiparametric logistics path model is reliable, especially when accuracy, sensitivity, specificity, and R^2_{total} support each other.

The results of this study can serve as a reference for state-owned enterprises in considering policies such as providing training on the use of organic pesticides and fertilizers as cost-effective alternatives for maintaining crops, ensuring price volatility remains stable while improving farmer quality. Additionally, it can also include support for entrepreneurship, such as compost trading, as a means of increasing farmers' income while waiting for harvest results.

REFERENCES

[1] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis*, Eighth Edition. Practice Hall, 2019.

[2] A. Agresti, *Categorical data analysis*, vol. 792. John Wiley & Sons, 2012.

[3] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, Second Edition. New York: Springer, 2008. doi: https://doi.org/10.1007/978-0-387-84858-7.

[4] W. Simon, *Generalized Additive Models: An Introduction with R, Second Edition*. New York: Chapman and Hall/CRC, 2017.

[5] A. Islamiyati, Anisa, M. Zakir, U. Sari, and D. S. Salam, "The Use Of The Binary Spline Logistic Regression Model on The Nutritional Status Data of Children," *Communications in Mathematical Biology and Neuroscience*, vol. 2023, 2023, doi: 10.28919/cmbn/7935.

[6] S. Arifin, "Pemodelan Regresi Logistik Ordinal dengan Estimator Spline Truncated," Universitas Hasanuddin, Makassar, 2023.

[7] M. F. Hidayat, R. F. Adji Achmad, and Solimun, "Estimation of Truncated Spline Function in Non-parametric Path Analysis Based on Weighted Least Square (WLS)," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jul. 2019. doi: 10.1088/1757-899X/546/5/052027.

[8] F. E. Harrell, "Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis Second Edition Springer Series in Statistics," USA, 2015. doi: 10.1007/978-3-319-19425-7.

[9] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, Fourth Edition. The Guilford Press, 2016.

[10] B. Vorley, M. Lundy, and J. MacGregor, "Business models that are inclusive of small farmers.," *Agro-industries for development*, pp. 186–222, 2009, doi: 10.1079/9781845935764.0186.

[11] Ardana, P. D. Heka, D. O. Suparwata, A. Sudrajat, S. Chatun, and I. Harsono, "The Role of Bali's Traditional Subak Farming System in the Preservation of Natural and Cultural Resources," *West Science Nature and Technology*, vol. 2, no. 01, pp. 31–38, 2024, doi: 10.58812/wsnt.v2i04.754.

[12] Coordinating Ministry for Economic Affairs of the Republic of Indonesia, "Government Strategies to Strengthen Food Security and Farmer Welfare." Accessed: Dec. 10, 2024. [Online]. Available: https://ekon.go.id/publikasi/detail/3044/strategi-pemerintah-mendorong-ketahanan-pangan-dan-kesejahteraan-petani

[13] D. Sahara *et al.*, "Sorghum Contribution to Increased Income and Welfare of Dryland Farmer Households in Wonogiri, Indonesia," *Agriculture*, vol. 13, no. 1609, 2023, doi: 10.3390/agriculture13081609.

[14] C. Martins-Filho and M. Toreto, *Relative Prices of Food and the Volatility of Agricultural Commodities: Evidence for a Panel of Developing Economies*, vol. 11. Springer, 2016. doi: 10.1007/978-3-319-28201-5_11.

[15] A. Oktavia, ; Zulfanetti, and ; Yulmardi, "Analisis produktivitas tenaga kerja sektor pertanian di Sumatera," *Jurnal Paradigma Ekonomika*, vol. 12, no. 2, pp. 2085–1960, 2017, doi: https://doi.org/10.22437/paradigma.v12i2.3940.

[16] Solimun, F. Adji, and L. Dwi, "Pengembangan Pemodelan Analisis Jalur Semiparametrik Smoothing Spline Dengan Pendekatan Sains Data: Variable And Relationship Mining Pada Inklusivitas Petani," 2023.

[17] C. I. Nyoman Budiantara, Ms. Bambang Widjanarko Otok, and Ms. Suhartono, "Spline Estimator in Birespon Nonparametric Regression for Longitudinal Data (Case Study Patient with Pulmonary Tuberculosis in Malang)," Institut Teknologi Sepuluh Nopember, Surabaya, 2016.

[18] C. R. Rao, *Linear Statistical Inference and Its Application*. John Wiley and Sons, 1965. doi: 10.1002/9780470316436.

[19] A. Agresti, *Foundations of linear and generalized linear models*, First Edition. John Wiley & Sons, 2015.

[20] Solimun, *Structural Equation Modelling (SEM) Lisrel dan Amos*. Malang: Fakultas MIPA Universitas Brawijaya, 2002.