Journal of Statistics and Its Application, Likelihood Vol 1 No 1, June 2025

Spline Truncated Nonparametric Regression Estimator on the Status of Unmet Need in East Java Province 2023

Afiqah Saffa Suriaslan[1], Sidratul Muthahharah[2], Jiran Julita[3], Muh Ilham Dirwan[4]

[1] Institut Teknologi Sepuluh Nopember, Indonesia, email: 6003231017@student.its.ac.id
[2] Institut Teknologi Sepuluh Nopember, Indonesia, email: sidrasssss11@gmail.com
[3] Institut Teknologi Sepuluh Nopember, Indonesia, email: jiranjulita11@gmail.com
[4] Universitas Airlangga, Indonesia, email: muh.ilham.dirwan-2023@fib.unair.ac.id

Abstract

Nonparametric regression has been widely developed for quantitative responses, yet methodological extensions for categorical response data remain relatively limited. In particular, although truncated spline estimators are well known for accommodating changing regression patterns through knot points, their application to categorical outcomes has not been adequately developed in previous studies. This study aimed to develop a truncated spline nonparametric regression estimator for categorical data and to evaluate its performance in modelling the status of unmet need in East Java Province in 2023. The study combined theoretical development with an empirical application using secondary data from 38 regencies/cities and three predictor variables related to demographic and socio-economic conditions. Model performance was assessed using deviance, classification accuracy, sensitivity, specificity, area under the curve (AUC), and Press's Q statistic. The results show that the proposed truncated spline nonparametric regression outperformed binary logistic regression, yielding a smaller deviance value (15.9745 vs. 30.9507), higher accuracy (86.8421% vs. 76.3158%), and stronger discriminatory ability, as indicated by a higher AUC (86.3077% vs. 72.7692%). The proposed model also produced a larger Press's Q statistic (20.6316), exceeding the chi-square benchmark and indicating classification performance better than chance. These findings suggest that truncated spline nonparametric regression provides a flexible and effective alternative for modelling categorical outcomes characterized by potentially nonlinear relationships.

Keywords: Categorical Data, Spline Truncated, Nonparametric Regression

Introduction

Spline Truncated nonparametric regression can be used to determine the relationship between the response and predictor variables when the function of the regression curve with changing patterns that depend on knot points. The nonparametric regression curve is assumed to be smooth in the sense that it is contained in a certain function space. The data were expected to find their own form of estimation, without being influenced by the subjective factors of the researcher. Thus, the nonparametric regression approach is highly flexible. The nonparametric approach can be implemented based on observed data using smoothing techniques. There are many smoothing techniques, including the Spline estimator (Eubank, 1998), Fourier Series estimator (Bilodeau, 1992), Wavelet estimator (Li, 2006), Kernel estimator (Okumura and Naito, 2006), and Local Polynomial (Sua and Ullah, 2008).

Spline estimators are used for data with changing patterns that depend on knot points (Eubank, 1998). A local polynomial estimator that has been used to reduce the bias properties and asymptotic variance of the local polynomial estimator in nonparametric regression with more than one response variable (Welsh and Yee, 2006). Wavelet estimator that has been used to model observations of a signal contaminated with noise, which is Gaussian distributed and

https://doi.org/10.54065/likelihood.564

additive (Antoniadis et.al., 2001). The Fourier Series estimator is used for patterned data that tend to repeat (Bilodeau, 1992). Among these estimators, the Fourier Series method was used to. This method is very specialized and well used in data cases in which the response and predictor variables exhibit a repeating pattern following a certain trend (Tripena and Budiantara, 2007). These smoothing approaches indicate that nonparametric regression has developed into a broad and adaptable framework for modelling complex functional relationships that cannot be adequately represented through rigid parametric assumptions. In practical applications, the choice of estimator depends strongly on the data pattern, the smoothness of the underlying regression function, and the behavior of the response variable. Therefore, selecting an appropriate estimator is not merely a technical exercise, but also a substantive modelling decision that determines how well the method can capture the actual structure of the observed phenomenon.

Some researchers have developed nonparametric regression estimators for categorical data, such as (Suliyanto et.al., 2020) using Local Likelihood Logit Estimation, (Hamie et.al., 2022) using the Decision Tree approach, and (Wang et.al., 2023) using the B-Spline function and recently, researchers have developed estimators for nonparametric regression using categorical data. However, these studies only used certain functions. No previous study has developed a nonparametric regression estimator for categorical response data using a Spline Truncated function. This condition reveals an important methodological gap in the current literature. Although nonparametric modelling has been widely extended for quantitative responses, the development of flexible estimators for categorical response variables remains relatively limited, especially when the objective is to accommodate regression curves with local structural changes. In many real-world applications, particularly in public health, demography, and socio-economic studies, the response variable is frequently categorical, while the relationship with predictors is nonlinear and difficult to specify a priori. Consequently, the absence of a Spline Truncated-based nonparametric estimator for categorical responses represents not only a technical limitation but also a missed opportunity to provide a more flexible modelling framework for complex binary or categorical outcomes.

It is not always the case that the plot follows a linear trend, which sometimes results in non-patterned and repeating patterns. As a result, a conventional nonparametric regression estimator (quantitative data) is not appropriate and tends to produce large errors (Zulfadhli et.al., 2024). In this study, a nonparametric regression estimator of Spline Truncated was developed with response variables in the form of categorical data. The need for such development becomes more relevant when the observed relationship between predictors and response exhibits heterogeneous local behavior, where changes in slope, turning points, or segmented patterns may emerge across different intervals of the predictor space. Under such conditions, estimators designed primarily for continuous responses may fail to provide accurate classification or probability estimation when applied directly to categorical outcomes. Therefore, extending the Spline Truncated framework to categorical response modelling is a logical and necessary step to improve flexibility, interpretability, and predictive performance in nonparametric regression settings.

From an applied perspective, this methodological development is particularly relevant for analysing the status of unmet need, which constitutes an important issue in population and reproductive health studies. In the context of East Java Province in 2023, unmet need remains a meaningful policy indicator because it reflects the extent to which women or couples who wish to delay or limit births are not adequately served by family planning programs. Such a condition is associated with demographic characteristics, socio-economic conditions, and access to health services, all of which may interact in nonlinear ways. When these relationships are modelled using overly restrictive assumptions, important local features in the data may be overlooked. For this reason, a more adaptive regression approach is required to capture the actual structure of the relationship between predictor variables and the categorical status of unmet need. The application presented in this study therefore provides a relevant empirical setting for evaluating whether the proposed Spline Truncated estimator can offer improved modelling performance compared with more conventional approaches.

The novelty of this study lies in the development of a Spline Truncated nonparametric regression estimator specifically for categorical response data. Unlike previous studies that employed local likelihood, decision tree, or B-spline-based approaches for categorical outcomes, this study introduces a truncated spline formulation that is designed to accommodate changes in curve patterns through knot points while maintaining the flexibility required for categorical-response modelling. This is an important extension because the truncated spline has been widely recognized as effective for modelling piecewise behavior in quantitative-response settings, yet its potential for categorical data has not been formally developed in previous studies. By bridging this gap, the present study contributes a new alternative estimator to the growing literature on nonparametric regression for qualitative outcomes.

The contribution of this study is threefold. First, it extends the methodological scope of nonparametric regression by proposing a new estimator for categorical response variables based on the Spline Truncated function. Second, it provides a comparative empirical assessment by evaluating the proposed estimator against binary logistic regression using deviance, AUC, and Press's Q as performance measures. Third, it demonstrates the practical relevance of the proposed approach through an application to unmet need status data in East Java Province, thereby linking methodological innovation with a substantive public-policy issue. In this sense, the study contributes not only to the statistical literature on flexible regression modelling but also to the broader field of evidence-based demographic and health analysis.

Accordingly, this study aims to develop a Spline Truncated nonparametric regression estimator for categorical response data and to assess its performance in modelling the status of unmet need in East Java Province in 2023. By doing so, this study is expected to provide both theoretical and empirical contributions. Theoretically, it enriches the development of nonparametric regression methods for categorical data. Empirically, it offers an alternative modelling strategy that may produce better classification performance and more accurate estimation in cases where the relationship between predictors and a categorical response is nonlinear, locally changing, and difficult to capture using conventional parametric methods.

Method

In applying the multivariable Spline Truncated nonparametric regression method for categorical data, we use application data the status of unmet need in East Java Province in 2023. The data used is secondary data sourced from dynamic tables on the BKKBN website as well as the Provincial publications in figures of East Java province. The data consists of 1 response variable (y) and 3 predictor variables (x). The variables are detailed in Table 1.

Table 1. Variable Description

Variable	Notation	Description	Unit	Scale
Response	y	Status of Unmet Need	0 = Target Achieved 1 = Target not Achieved	Nominal
Predictor	x_1	Percentage of Family Heads with no Primary Education	Percent	Rasio
	x_2	Percentage of Couples of Childbearing Age (PUS) with 2 Children	Percent	Rasio
	x_3	Percentage of Couples of Childbearing Age (PUS) who Seek Services at FKTP (Health Centers or Equivalent, Doctor's Practices, Private Clinics or Equivalent and Class D private Hospitals or Equivalent)	Percent	Rasio

38 district exist in East Java and shown in Table 2 below.

Status	District	Total
Target Achieved ($y = 0$)	Pacitan, Trenggalek, Blitar, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Ngawi, Bojonegoro, Tuban, Lamongan, Pamekasan, Kota Kediri, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun	24 Districts
Target Not Achieved ($y = 1$)	Ponorogo, Tulungagung, Kediri, Malang, Sidoarjo, Magetan, Gresik, Bangkalan, Sampang, Sumenep, Kota Blitar, Kota Malang, Kota Surabaya, Kota Batu	14 Districts

Based on Table 2, there is still a lot of Unmet Need, especially in East Java 2023.

Results and Discussion

Descriptive Analytics

Descriptive analysis is used to determine the characteristics of the data for each variable as follows Table 3.

Table 3. Descriptive Statistics of Research Variables

Variable	Mean	Variance	Min	Max
y	-	-	-	-
x_1	8.38	-0.41	5.07	11.82
x_2	7.06	-0.78	2.38	54.11
x_3	84.59	-2.65	50.30	98.18

Table 3 provides information about the characteristics of the variables, which is the status of unmet need in 38 regencies/cities in East Java 2023. In addition, it is obtained that each variable does not have missing values and there is no multicollinearity between predictor variables. The conceptual predictor variable used in this study is as follows in figure 1.

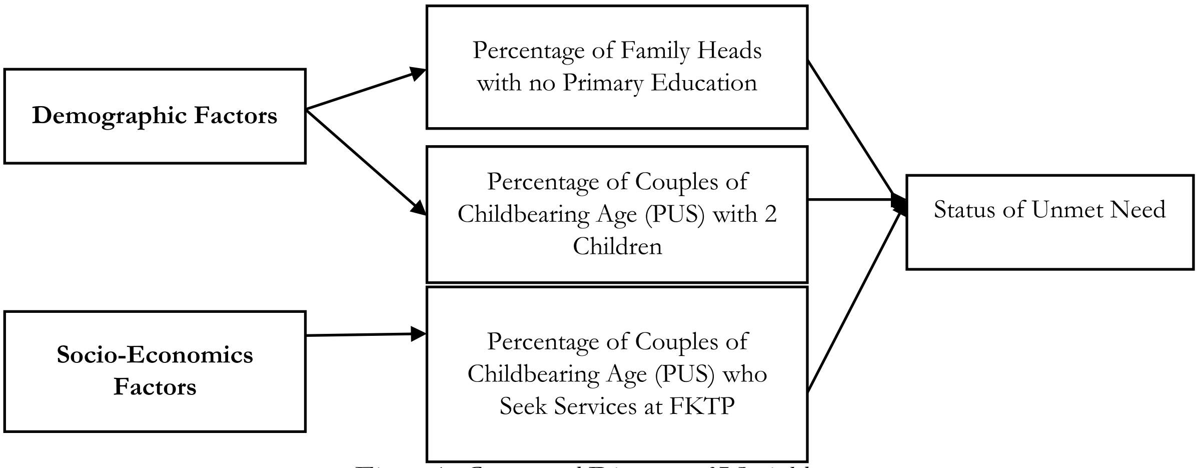

Figure 1. Conceptual Diagram of Variables

Based on Figure 1, unmet need is a problem that covers many aspects including demographic factors and socio-economics factors. Most women of childbearing age who want to stop or adjust their spacing are unable to use family planning methods. Demographers and health experts refer to these women of childbearing age as having an unmet need for family planning services (Kumar and Singh, 2013).

The Spline Truncated nonparametric regression model is as follows.

$$\pi(\boldsymbol{x_i}) = \frac{e^{\beta_0 + \sum_{j=1}^{p}\left(\sum_{k=0}^{m} \beta_{jk} x_{ji}^{k} + \sum_{u=1}^{r} \beta_{j(m+u)}\left(x_{ji} - K_{ju}\right)_+^m\right)}}{1+e^{\beta_0 + \sum_{j=1}^{p}\left(\sum_{k=1}^{m} \beta_{jk} x_{ji}^{k} + \sum_{u=1}^{r} \beta_{j(m+u)}\left(x_{ji} - K_{ju}\right)_+^m\right)}} \; ; \; i = 1,2,\ldots,n \quad (1)$$

Where K_j is knot point, m is degree, u is the number of knot and p is the number of predictor variables. Truncated function for Spline is defined as follows:

$$\left(x_{ji} - K_{ju}\right)_+^m = \begin{cases} \left(x_{ji} - K_{ju}\right)^m & , x_{ji} \geq K_{ju} \\ 0 & , x_{ji} < K_{ju} \end{cases}$$

Parameter Estimation

Based on Spline Truncated nonparametric regression model (1), parameter estimation results on the Spline Truncated nonparametric regression model with the optimal number of knot combination $x_1 = 1, x_2 = 2, x_3 = 2$ and 1 degree for data on the status of unmet need in East Java in 2023 are as follows.

$$\hat{\pi}(\boldsymbol{x_i}) =$$

$$\frac{e^{40.84 - 3.50x_{1i} + 4.23(x_{1i} - 2.11)_+^1 - 1.87x_{2i} + 3.82(x_{2i} - 31.63)_+^1 - 180.15(x_{2i} - 34.92)_+^1 + 0.28x_{3i} - 1.27(x_{3i} - 74.23)_+^1 - 0.20(x_{3i} - 83.34)_+^1}}{1 + e^{40.84 - 3.50x_{1i} + 4.23(x_{1i} - 2.11)_+^1 - 1.87x_{2i} + 3.82(x_{2i} - 31.63)_+^1 - 180.15(x_{2i} - 34.92)_+^1 + 0.28x_{3i} - 1.27(x_{3i} - 74.23)_+^1 - 0.20(x_{3i} - 83.34)_+^1}}$$

This section is also a major part of the research articles and is also usually the longest part of an article. Discussion of the research presented in this section are the result. The process of data analysis such as statistical calculations or other processes for the achievement of its research. Please present the discussion narratively.

More details can be seen in Table 4.

Table 4. Parameter Estimation

Parameters	Estimations
β_0	40.836567
β_{11}	-3.5020957
β_{12}	4.2345235
β_{21}	-1.8746554
β_{22}	3.8197080
β_{23}	-180.1522119
β_{31}	0.2836886
β_{32}	-1.2729495
β_{33}	-0.2022823

Comparison of Binary Logistic Regression and Spline Truncated Nonparametric Regression

Based on Spline Truncated nonparametric regression model (1), parameter estimation results on the Spline Truncated nonparametric regression model with the optimal number of knot combination $x_1 = 1, x_2 = 2, x_3 = 2$ and 1 degree for data on the status of unmet need in East Java in 2023 are as follows.

Table 5. Comparison of Deviance Values

Methods	Deviance Values
Binary Logistic Regression	30.9507
Spline Truncated Nonparametric	15.9745

Based on Table 5, the deviance value for the Spline Truncated nonparametric regression (15.9745) was smaller than binary logistic regression (30.9507).

The selected Spline Truncated nonparametric regression model demonstrated the highest AUC or the smallest Press's Q. Using the classification test, the following results are obtained in Table 6.

Table 6. Comparison of AUC and Press's Q

Methods	Accuracy	Sensitivity	Specificity	AUC	Press's Q value	Chi Square
Binary Logistic Regression	76.3158	84	61.5385	72.7692	10.5263	3.8415
Spline Truncated Nonparametric Regression	86.8421	88	84.6154	86.3077	20.6316	3.8415

Based on Table 6, shows that the AUC value of Spline Truncated nonparametric regression (86.30%) is higher than binary logistic regression (72.76%). In addition, a larger Press's Q value for Spline Truncated nonparametric regression (20.6316).

Conclusion

The results of this study indicate that the Spline Truncated nonparametric regression model outperformed the binary logistic regression model in analysing the status of unmet need in East Java Province in 2023. This superiority is reflected in its smaller deviance value of 15.9745, compared with 30.9507 for binary logistic regression, indicating a better model fit to the observed data. In terms of classification performance, the Spline Truncated model also produced higher predictive accuracy, with an accuracy of 86.8421%, sensitivity of 88%, specificity of 84.6154%, and an AUC of 86.3077%, whereas the binary logistic regression model achieved a lower AUC of 72.7692%. Furthermore, the Press's Q value for the Spline Truncated model was 20.6316, which exceeded both the corresponding value for binary logistic regression (10.5263) and the chi-square benchmark (3.8415), indicating that the resulting classification was statistically better than chance. These findings suggest that the proposed Spline Truncated nonparametric regression provides a more flexible and effective approach for modelling categorical unmet need data characterized by potentially nonlinear patterns. Therefore, this model may serve as a useful alternative to conventional binary logistic regression for similar classification problems in demographic and public health studies.

Acknowledgment

-

References

Ahinkorah, B. O., Ameyaw, E. K., & Seidu, A. A. (2020). Socio-economic and demographic predictors of unmet need for contraception among young women in sub-Saharan Africa: Evidence from cross-sectional surveys. *Reproductive Health, 17*, 1-11. https://doi.org/10.1186/s12978-020-0895-9

Antoniadis, A., Bigot, J., & Spatinas, T. (2001). Wavelet estimators in nonparametric regression: A comparative simulation study. *Journal of Statistical Software, 6*, 1-83. https://doi.org/10.18637/jss.v006.i01

Bilodeau, M. (1992). Fourier smoother and additive models. *The Canadian Journal of Statistics, 3*, 257-259. https://doi.org/10.2307/3315115

Eubank, R. L. (1998). *Spline smoothing and nonparametric regression*. Marcel Dekker.

Hamie, H., Hoayek, A., El-Ghoul, B., & Khalifeh, M. (2022). Application of non-parametric statistical methods to predict pumpability of geopolymers for well cementing. *Journal of Petroleum Science and Engineering, 212*, 110333. https://doi.org/10.1016/j.petrol.2022.110333

Kumar, A., & Singh, A. (2013). Trends and determinants of unmet need for family planning in Bihar (India): Evidence from National Family Health Surveys. *Advances in Applied Sociology, 3*(2), 157–163. https://doi.org/10.4236/aasoci.2013.32022

Li, L. (2002). Nonlinear wavelet-based nonparametric curve estimation with censored data and inference on long memory processes. ProQuest Information and Learning Company.

Zulfadhli, M., Budiantara, I. N., & Ratnasari, V. (2024). Nonparametric regression estimator of multivariable Fourier series for categorical data. *MethodsX, 11*, 102983. https://doi.org/10.1016/j.mex.2023.102983

Okumura, H., & Naito, K. (2006). Non-parametric kernel regression for multinomial data. *Journal of Multivariate Analysis, 97*, 2009-2022. https://doi.org/10.1016/j.jmva.2005.10.002

Sua, L., & Ullah, A. (2008). Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics, 144*, 193-218. https://doi.org/10.1016/j.jeconom.2007.10.004

Suliyanto, S., Rifada, M., & Tjahjono, E. (2020). Estimation of nonparametric binary logistic regression model with local likelihood logit estimation method (Case study of diabetes mellitus patients at Surabaya Hajj General Hospital). In *Symposium on Biomathematics 2019* (pp. 1551-7616). AIP Conference Proceedings, Bali. https://doi.org/10.1063/1.5145473

Tripena, A., & Budiantara, I. N. (2007). Fourier estimator in nonparametric regression. In *International Conference on Natural Sciences and Applied Natural Sciences* (pp. 2-4). Yogyakarta.

Welsh, A. H., & Yee, T. W. (2006). Local regression for vector responses. *Journal of Statistical Planning and Inference, 136*, 3007–3031. https://doi.org/10.1016/j.jspi.2005.12.007

Wang, T., Tang, W., Lin, Y., & Su, W. (2023). Semi-supervised inference for nonparametric logistic regression. *Statistics in Medicine, 42*, 2573–2589. https://doi.org/10.1002/sim.9699