For example, a scatter plot between the response variable and each predictor variable with a sample size of $n = 50$ and $\sigma^2 = 0.05$ shown in Fig. 1.

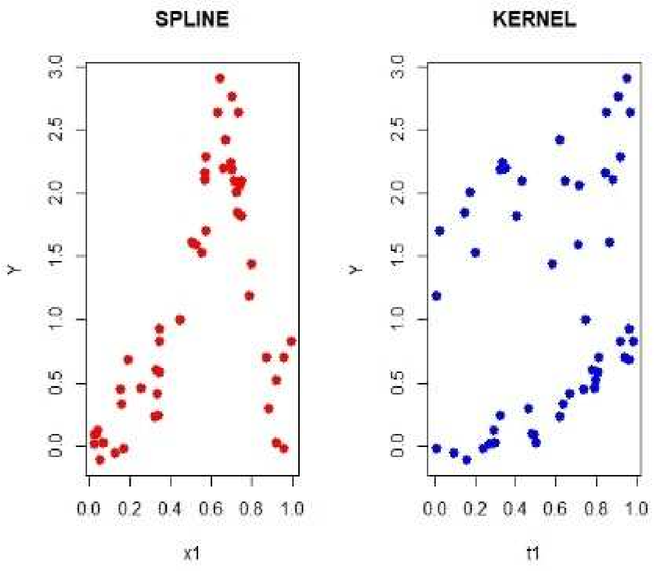


Fig. 1 Scatter plot between predictor and response variable with *n=50*

Furthermore, Fig. 2 shows the scatter plot with a sample size of $n = 200$ and $\sigma^2 = 0.05$.

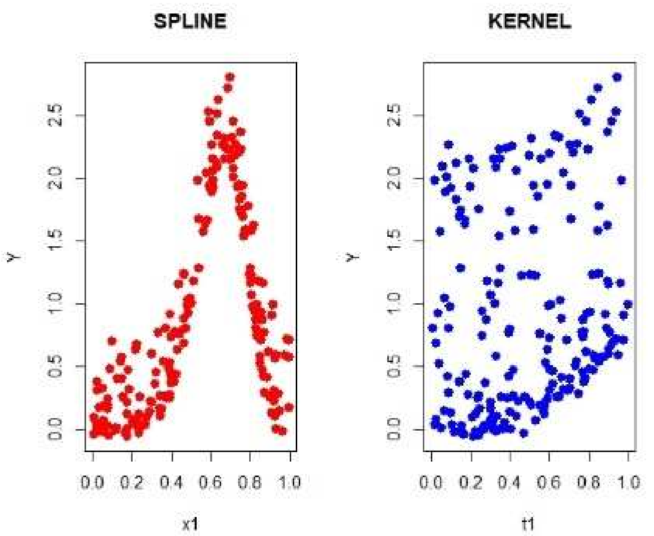


Fig. 2 Scatter plot between predictor and response variable with *n=200*

Fig. 1 and Fig. 2 show that each predictor variable has a different form of regression curve. Variable x_1 shows the characteristics of the Truncated Spline estimator, which has a changing data pattern at certain sub-intervals. In comparison, variable t_1 shows a data pattern that does not have a certain pattern, so that it was modeled with the Kernel estimator. Furthermore, based on the scatter plot in Fig. 1 and Fig. 2, a nonparametric regression model was applied using a mixed estimator of the Truncated Spline and Gaussian Kernel.

The number of knot points to be tested is only one-knot point for variables defined as a Truncated Spline component. The simulation results in the form of the average CV, GCV, UBR, and coefficient of determination (R^2) are presented in Tables 1, 2, and 3.

TABLE I
SIMULATION RESULTS WITH THE CV METHOD

Variance	Average	Number of Samples			
		n=25	*n=50*	*n=100*	*n=200*
$\sigma^2 = 0.05$	CV	0.136	0.120	0.110	0.109
	R^2	85.65%	84.75%	84.33%	84.45%
$\sigma^2 = 0.5$	CV	0.393	0.374	0.371	0.364
	R^2	66.38%	65.12%	61.84%	61.02%
$\sigma^2 = 1$	CV	1.128	1.133	1.133	1.087
	R^2	40.86%	37.42%	38.05%	35.41%

TABLE II
SIMULATION RESULTS WITH THE GCV METHOD

Variance	Average	Number of Samples			
		n=25	*n=50*	*n=100*	*n=200*
$\sigma^2 = 0.05$	GCV	1.473	2.733	5.010	8.403
	R^2	86.48%	85.30%	85.26%	85.34%
$\sigma^2 = 0.5$	GCV	1.475	3.197	4.436	10.153
	R^2	71.26%	67.25%	63.38%	61.89%
$\sigma^2 = 1$	GCV	2.168	4.033	8.096	15.192
	R^2	56.70%	44.86%	41.18%	37.15%

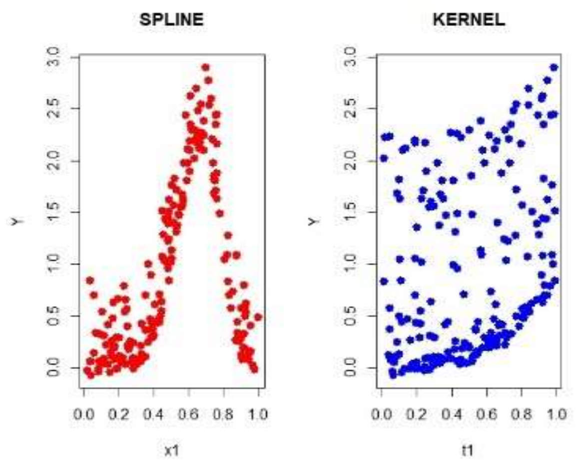
TABLE III
SIMULATION RESULTS WITH THE UBR METHOD

Variance	Average	Number of Samples			
		n=25	*n=50*	*n=100*	*n=200*
$\sigma^2 = 0.05$	UBR	0.009	0.008	0.004	0.004
	R^2	83.71%	82.74%	82.72%	82.59%
$\sigma^2 = 0.5$	UBR	0.014	0.007	0.005	0.004
	R^2	64.23%	62.38%	59.42%	59.78%
$\sigma^2 = 1$	UBR	0.016	0.009	0.006	0.004
	R^2	38.28%	35.75%	36.19%	34.04%

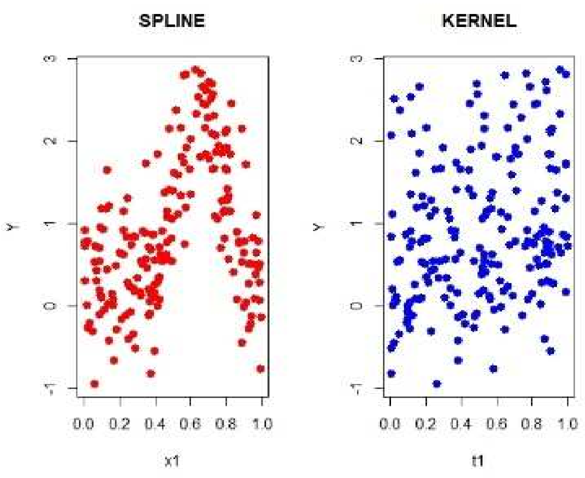
For the various sample sizes *n*, such as 25, 50, 100, and 200, with all variations of the variance tested, the GCV method provides better knot point and bandwidth estimation results compared to the CV and UBR methods. This is indicated by the value of the coefficient of determination (R^2) obtained from each experiment with GCV, which is higher than the other two methods. Furthermore, the residuals of each modeling results for each combination of sample size variation and variance follow a normal distribution.

For example, the number of samples *n=25* and the error variance is , using the GCV method in selecting the optimal knot point and bandwidth, the average GCV value is 1.473 with an R2 value of 86.48%. Meanwhile, using the CV method and the same conditions obtained an average CV value is 0.136 with R^2 value is 85.65%. Using the UBR method, it was obtained an average UBR value of 0.009 and R^2 value is 83.71%.

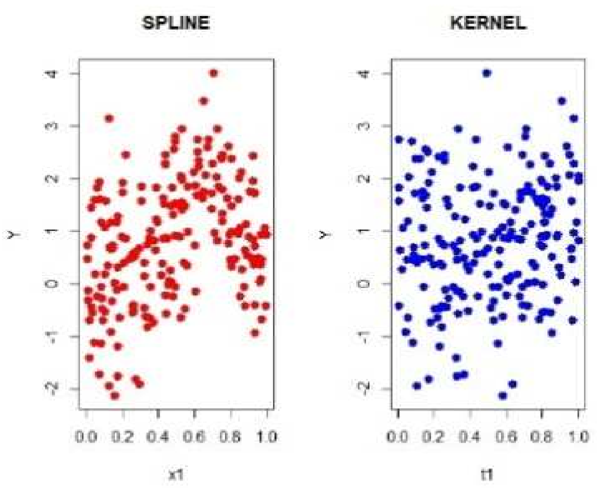
The impact of the variation variance measures σ^2 in this study has an effect on the simulation results. It can be seen that the increase of the variance tested, the value of R^2 for all methods used both CV, GCV, and UBR tend to decrease. The variance shows the deviation of the data from the average, so that the higher the variance value that is tried, then there was a tendency for the data spread far from the average value. The illustration of generated data with *n=200* and various variance conditions are shown in Fig. 3.



(a)



(b)



(c)

Fig. 3 Illustration of impact by variance (a) $\sigma^2 = 0.05$; (b) $\sigma^2 = 0.5$; (c) $\sigma^2 = 1$

Based on Fig. 3, Thus, it can be concluded that the size of the sample tested and the variance size is important. Moreover, it can be seen that in the variance $\sigma^2 = 0.05$, the Truncated Spline component has clearly shown a changing pattern at certain sub-intervals. While the Gaussian Kernel component does not appear to have a certain pattern. The increasing of the variance value, for example $\sigma^2 = 1$, the data pattern for Truncated Spline component implicitly still has shown a changing pattern in certain sub-intervals, but there is a tendency for the pattern to spread. While the Gaussian Kernel component looks more spread out and doesn't have a pattern. Based on the impact of the variance size and sample size, it can be seen that the GCV method still gives the correct estimation of knot point and bandwidth, so it can provide better coefficient of determination (R^2) value compared to the other two methods for each condition.

Based on the simulation results, the knot point and bandwidth estimation results from the CV, GCV, and UBR methods are quite good. However, the GCV method provides better performance and accuracy for each combination of sample sizes and variance variations tried. The GCV method produces optimal knot point and bandwidth to obtain the largest coefficient of determination (R^2) for each combination. As a result, the GCV method is more suitable for estimating the knot point and bandwidth in the nonparametric regression model mixed estimator of the Truncated Spline and Gaussian Kernel.

IV. Conclusion

Simulation studies on the nonparametric regression model mixed estimator of the Truncated Spline and Gaussian Kernel to compare the performance of the Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) methods in estimating the optimal smoothing parameter (knot point and bandwidth) have been successfully carried out. Based on the simulation results, with an error following the Normal distribution and in a combination of sample size variation and error variance. The GCV method provides better result performance and accuracy than the CV and UBR methods. The GCV method produces optimal smoothing parameters so that the largest coefficient of determination (R^2) is obtained for each combination. The results obtained in this study have the potential to contribute to the development of statistics, especially in the field of nonparametric regression.

Acknowledgment

The authors gratefully acknowledge for the financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

References

[1] A. T. R. Dani, V. Ratnasari, and I. N. Budiantara, "Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, no. 1, p. 012020, 2021, doi: 10.1088/1757-899x/1115/1/012020.

[2] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker, 1999.

[3] I. N. Budiantara, V. Ratnasari, M. Ratna, and I. Zain, "The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties," *Appl. Math. Sci.*, vol. 9, no. 122, pp. 6083–6094, 2015, doi: 10.12988/ams.2015.58517.

[4] N. Y. Adrianingsih, I. N. Budiantara, and J. D. T. Purnomo, "Modeling with Mixed Kernel, Spline Truncated and Fourier Series on Human Development Index in East Java," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, no. 1, p. 012024, 2021, doi: 10.1088/1757-899x/1115/1/012024.

[5] D. R. Sari Saputro, K. R. Demu, and P. Widyaningsih, "Nonparametric truncated spline regression model on the data of human development index (HDI) in indonesia," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, pp. 6–10, 2018, doi: 10.1088/1742-6596/1028/1/012219.

[6] N. Chamidah, B. Lestari, A. Massaid, and T. Saifudin, "Estimating mean arterial pressure affected by stress scores using Spline Nonparametric Regression model approach," *Commun. Math. Biol. Neurosci.*, vol. 2020, pp. 1–12, 2020.

[7] G. Wahba, *Spline Models for Observational Data*, Pennsylvania: SIAM, 1990.

[8] N. P. A. M. Mariati, N. Budiantara, and V. Ratnasari, "Truncated Spline Estimation of Percentage Poverty Modeling in Papua Province," *ICSA - Int. Conf. Stat. Anal. 2019*, vol. 1, pp. 69–82, 2021, doi: 10.29244/icsa.2019.pp69-82.

[9] B. Fatmawati, Budiantara, I N; Lestari, "Comparison of Smoothing and Truncated Spline Estimators in Estimating Blood Pressure Models," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 685–707, 2019.

[10] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, "Smoothing Spline Estimator in Nonparametric Regression (Application: Poverty in Papua Province)," *Proc. 7th Int. Conf. Res. Implementation, Educ. Math. Sci. (ICRIEMS 2020)*, vol. 528, no. Icriems 2020, pp. 309–314, 2021, doi: 10.2991/assehr.k.210305.044.

[11] B. Lestari, Fatmawati, and I. N. Budiantara, "Spline estimator and its asymptotic properties in multiresponse nonparametric regression model," *Songklanakarin J. Sci. Technol.*, vol. 42, no. 3, pp. 533–548, 2020, doi: 10.14456/sjst-psu.2020.68.

[12] L. R. Cheruiyot, "Local linear regression estimator on the boundary correction in nonparametric regression estimation," *J. Stat. Theory Appl.*, vol. 19, no. 3, pp. 460–471, 2020, doi: 10.2991/jsta.d.201016.001.

[13] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "An extended model of penalized spline with the addition of Kernel Functions in nonparametric regression model," *Appl. Math. Inf. Sci.*, vol. 13, no. 3, pp. 453–460, 2019, doi: 10.18576/amis/130318.

[14] F. Yan, Q. S. Xu, M. L. Tang, and Z. Chen, "Kernel density-based likelihood ratio tests for linear regression models," *Stat. Med.*, vol. 40, no. 1, pp. 119–132, 2021, doi: 10.1002/sim.8765.

[15] N. Chamidah and T. Saifudin, "Estimation of children growth curve based on kernel smoothing in multi-response nonparametric regression," *Appl. Math. Sci.*, vol. 7, no. 37–40, pp. 1839–1847, 2013, doi: 10.12988/ams.2013.13168.

[16] I. Wayan Sudiarsa, "Simulations Study Combined Estimator Fourier Series and Spline Truncated in Multivariable Nonparametric Regression," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052074.

[17] D. R. S. Saputro, A. Sukmayanti, and P. Widyaningsih, "The nonparametric regression model using Fourier series approximation and penalized least squares (PLS) (case on data proverty in East Java)," *J. Phys. Conf. Ser.*, vol. 1188, no. 1, 2019, doi: 10.1088/1742-6596/1188/1/012019.

[18] A. Prahutama, Suparti, and T. W. Utami, "Modelling fourier regression for time series data - A case study: Modelling inflation in foods sector in Indonesia," *J. Phys. Conf. Ser.*, vol. 974, no. 1, pp. 0–9, 2018, doi: 10.1088/1742-6596/974/1/012067.

[19] M. F. F. Mardianto, S. M. Ulyah, and E. Tjahjono, "Prediction of national strategic commodities production based on multi-Response nonparametric regression with fourier series estimator," *Int. J. Innov. Creat. Chang.*, vol. 5, no. 3, pp. 1151–1176, 2019.

[20] I. Nyoman Budiantara *et al.*, "Modeling percentage of poor people in Indonesia using kernel and Fourier series mixed estimator in nonparametric regression," *Investig. Operacional*, vol. 40, no. 4, pp. 538–550, 2019, doi: 10.5281/zenodo.3721293.

[21] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "A reproducing kernel hilbert space approach and smoothing parameters selection in spline-kernel regression," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 2, pp. 465–475, 2019.

[22] R. Hidayat, I. N. Budiantara, B. W. Otok, and V. Ratnasari, "Kernel-Spline Estimation of Additive Nonparametric Regression Model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052028.

[23] P. Dewanti, I. Nyoman Budiantara, and A. T. Rumiati, "Modelling of SDG's Achievement in East Java Using Bi-responses Nonparametric Regression with Mixed Estimator Spline Truncated and Kernel," *J. Phys. Conf. Ser.*, vol. 1562, no. 1, 2020, doi: 10.1088/1742-6596/1562/1/012016.

[24] D. P. Rahmawati, I. N. Budiantara, D. D. Prastyo, and M. A. D. Octavanny, "Mixed Spline Smoothing and Kernel Estimator in Biresponse Nonparametric Regression," *Int. J. Math. Math. Sci.*, vol. 2021, 2021, doi: 10.1155/2021/6611084.

[25] P. Craven and G. Wahba, "Smoothing noisy data with spline functions - Estimating the correct degree of smoothing by the method of generalized cross-validation," *Numer. Math.*, vol. 31, no. 4, pp. 377–403, 1978, doi: 10.1007/BF01404567.

[26] Y. Wang, "Smoothing spline models with correlated random errors," *J. Am. Stat. Assoc.*, vol. 93, no. 441, pp. 341–348, 1998, doi: 10.1080/01621459.1998.10474115.

[27] H. Nurcahayani, I. N. Budiantara, and I. Zain, "The Curve Estimation of Combined Truncated Spline and Fourier Series Estimators for Multiresponse Nonparametric Regression," *Mathematics*, vol. 9, no. 10, p. 1141, 2021.

[28] A. R. Devi, I. N. Budiantara, and V. Ratnasari, "Unbiased risk and cross-validation method for selecting optimal knots in multivariable nonparametric regression spline truncated (case study: Unemployment rate in Central Java, Indonesia, 2015)," *AIP Conf. Proc.*, vol. 2021, no. December, 2018, doi: 10.1063/1.5062767.

[29] T. W. Utami, M. A. Haris, A. Prahutama, and E. A. Purnomo, "Optimal knot selection in spline regression using unbiased risk and generalized cross validation methods," *J. Phys. Conf. Ser.*, vol. 1446, no. 1, 2020, doi: 10.1088/1742-6596/1446/1/012049.

[30] B. A. M. Al-Talib and A. A. Hammodat, "Using Some Wavelet Shrinkage Techniques and Robust Methods to Estimate the Generalized Additive Model Parameters in Non-Linear Models," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, p. 2344, 2020, doi: 10.18517/ijaseit.10.6.12767.