BIU Journal of Basic and Applied Sciences 10(1): 79 – 86, 2025.

ISSN: 2563-6424

REGRESSION SPLINES: EXTENDING POLYNOMIAL REGRESSION AND PIECEWISE CONSTANT FUNCTION MODELS

***AGU, S. C.[1] AND ELUGWU, F.[2]**
[1]Department of Computer Science, Benson Idahosa University, Benin City, Nigeria.
[2]Department of Computer Science, Delta State Polytechnic, Otefe-Oghara, Nigeria
*Corresponding author: sagu@biu.edu.ng

ABSTRACT
Polynomial regression models have the limitation of forcing a common construction on the non-linear function of X while the piecewise-constant function models suffer from over fitting a model unless there are natural breakpoints in the predictors. In view of these, the objective of this study considered how the polynomial regression and piecewise constant regression were extended by regression spline techniques to resolve their inadequacies. The study revealed that amongst the three regression spline methods: First, the piecewise polynomial attempted to enhance the hitches, but it fitted a discontinuous curve model which is a big problem. Second, although the cubic spline fitted a continuous curve, it gave high variance towards the extreme values of X. And third, the natural spline yielded a more stable curved model whose variance and the corresponding confident intervals are narrower near the ends of X. The study also highlighted how each of the regression spline methods hierarchically resolved the setbacks of the prior approach.

KEYWORDS: *Piecewise Polynomial regression, Cubic spline, Natural spline, Continuity Function, Degree of Freedom, Non-linear functions*

INTRODUCTION

The piecewise polynomial, the cubic spline, and the natural splines are the three fundamental regression spline algorithms (Gareth *et al.*, 2017) which are commonly used for piecewise interpolation of data (Hernandez, 2022). Particularly, cubic splines provide a smooth, continuous transition between the different sections or segments determined by the nodes or “knots” of the spline.

Polynomial regression has the problem of representing a given function globally, where the use of a single high-degree polynomial will produce large errors away from the central region (John, 2003). (Roksolana, 2019) showed that it either over-predicts or under-predicts a model fit and in curbing this limitation, a piecewise polynomial divides the domain of the function into segments where the function is represented by different polynomials in each segment, constraining the polynomials to at least have the same value at the segment boundaries, noting that the points where the polynomials meet are called knots. According to (Carl, 2002), a piecewise polynomial of order k with break sequence ξ (necessarily strictly increasing) is a function f that, on each of the half-open intervals $[\xi_j \,..\, \xi_{j+1}]$, agrees with some polynomial of degree $< k$, which makes the function right-continuous.

Considering the bodies of research in this regard, an accurate and efficient stability prediction of machining parameters was determined by a cubic Hermite–Newton approximation method which adopted piecewise polynomial interpolation that utilized two adjacent time intervals to correct the prediction (Yi *et al.*, 2002). According to Song and Lee (2007), piecewise polynomial method was proposed as an optimal approximation algorithm in time sensitive applications and in embedded systems with limited resources where the runtime of the approximate function is as crucial as its accuracy. The method also emerged as the outcome of the four families of low-degree polynomial algorithms to approximate curvilinear profiles extracted from 3D point clouds representing real objects (Chiara *et al.*, 2023). Mathematical programming formulations for solving optimization problems with piecewise polynomial (PWP) constraints were compared by globally solving six test sets of MINLPs and a realistic petroleum production optimization problem. Interestingly, the proposed framework showed promising numerical performance and facilitates the solution of PWP-constrained optimization problems using standard MINLP software (Bjarne and Brage, 2020). Regressions splines with survival time for cancer patients were presented in real-life clinical data which revealed that the quadratic spline yielded the most appropriate modeling tool when the data between all knot locations and endpoints having a quadratic data structure is continuous and has different slopes (Dare and Adeleke, 2021). Employing adaptive step-size one-at-a-time (OAT) optimization method, (Hernandez, 2022) utilized spline regression for filtering/extracting noise from data and for obtaining a smoother interpolation function, and cubic spline regression model was applied to determine the effect of four influencing indicators to Human Development Index (HDI) in Indonesia with GCV value = 0.003539 and 3 optimal knot points for each independent variable (Mutia *et al.*, 2022). According to Rohit *et al.* (2021), five piecewise interpolating polynomials used for modeling lung field region in 2D chest X-ray images revealed that PCHIP interpolation method outperformed others with an execution time of 5.04873s. Predicting the risk of having bipolar amongst patients using cubic spline revealed how bipolar disorder build up slowly and lingers in the body for long without been noticed due to fluctuations in risk tendency of the mood scores (Ogoke *et al.*, 2016). And (Elhakeem *et al.*, 2022) demonstrated that the natural cubic splines partook in describing nonlinear growth trajectories and analyzed repeated bone mass measures to characterize bone growth patterns amongst individuals.

Piecewise Polynomial

Piecewise polynomial is a method of building distinct low-degree polynomials across different sections of X, in place of constructing a high-degree polynomial over the entire section of X. Taking an instance, a piecewise cubic polynomial operates by building a cubic regression model below (Gareth *et al.*, 2017).

$$y_i = m_0 + m_1 x_i + m_2 x_i^2 + m_3 x_i^3 + \epsilon_i \qquad (1)$$

the constants m_0, m_1, m_2, and m_3 vary in various portions of the section of X, and the joints where the constants adjust are called knots.
A piecewise cubic polynomial with one knot at a joint j is shown below

$$y_i = \begin{cases} m_{01} + m_{11}x_i + m_{21}x_i^2 + m_{31}x_i^3 + \epsilon_i & if\ x_i < j; \\ m_{02} + m_{12}x_i + m_{22}x_i^2 + m_{32}x_i^3 + \epsilon_i & if\ x_i \geq j. \end{cases}$$

Remark I: For a piecewise cubic polynomial with 1 knot, two dissimilar polynomial functions are fitted, one on the subset of the observations with $x_i < j$, and one on the subset of the observations with $x_i \geq j$. The first polynomial function has constants $m_{01}, m_{11}, m_{21}, m_{31}$, while the second has constants $m_{02}, m_{12}, m_{22}, m_{32}$, and least squares can be applied in fitting each of these polynomial functions.
Remark II: Having T different knots throughout the range of X allows fitting $T + 1$ distinct cubic polynomials, and using more knots leads to a more flexible piecewise polynomial
Remark III: It is not recommended to use a cubic polynomial, instead, we fit piecewise linear functions. Indeed, the piecewise constant functions is a piecewise polynomial of degree 0 (Hernandez, 2022).
Figure 1 shows a piecewise cubic polynomial fit to a subset of the Wage dataset with a single knot at $age = 50$ (Gareth *et al.*, 2017). There is a problem in the model fit: the function is discontinuous and appears absurd. This is because each polynomial has four parameters. As a result, a total of 8 degrees of freedom were used in fitting the piecewise cubic polynomial model. This problem is rectified by using a cubic spline.

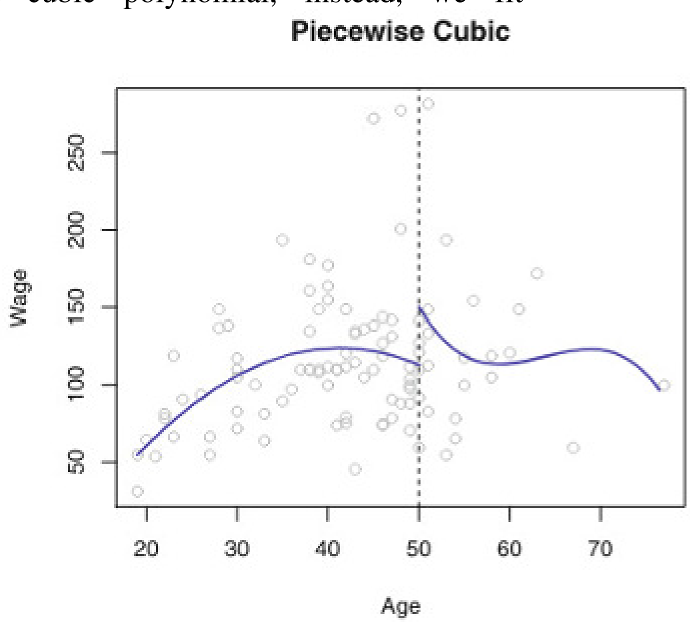

Fig. 1: Cubic polynomial at knot = 50 without any constraint

Cubic Spline

A cubic spline is a piecewise interpolation model that fits a cubic polynomial to each piece in a piecewise function (Brendan, 2022). At every point where two polynomials join, the first and the second derivatives are equal which results to a smooth fitting line. Cubic spline resolved the problem in the discontinuous fitted curves of the piecewise cubic polynomial by constraining the fitted curve to be continuous (employing continuity function). That is, the fitted line is restricted from jumping when $age = 50$. Figure 2 shows the resulting fit (Gareth *et al*., 2017). Although there is an improvement in the curve, the V-shaped join seems abnormal and needs to be corrected by making it smooth. To achieve this, two extra constraints were added by computing the first and second derivatives to the piecewise polynomial to be continuous when $age = 50$ as shown in Figure 3.

Remark IV: Each enforced constraint on the piecewise cubic polynomials successfully frees up one degree of freedom, by reducing the complexity of the resulting piecewise polynomial fit.

Based on Remark IV, the three constraints in Figure 3 (continuity, continuity of the first derivative, and continuity of the second derivative) were enforced which reduced the 8 degrees of freedom in Figure 1 to 5 degrees of freedom (Gareth *et al*., 2017), (Brendan, 2022).

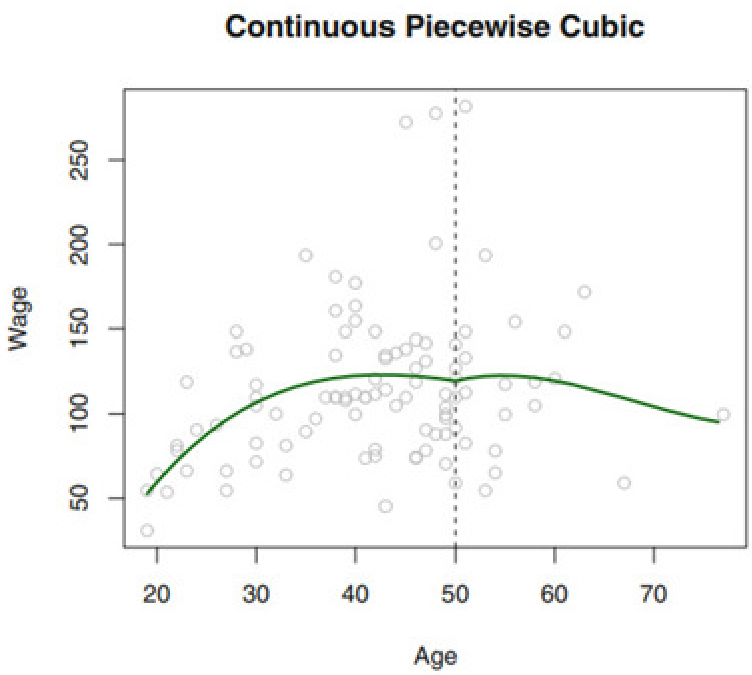

Fig. 2: Cubic polynomial at knot = 50 constrained to be continuous

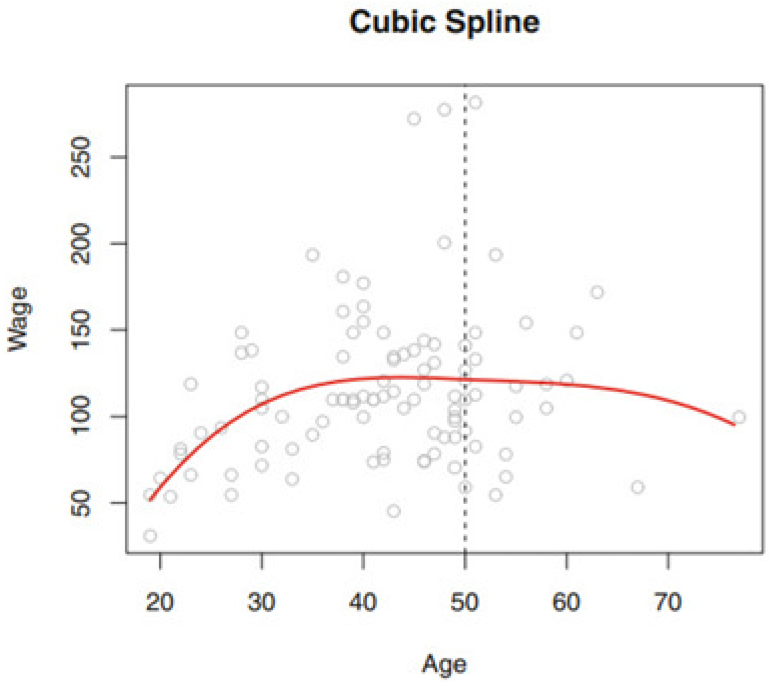

Fig. 3: Cubic polynomial at knot = 50 with 3 constraints – continuous, continuous first derivative, and continuous second derivative

Remark V: The curve in Figure 3 is called a cubic spline. Broadly speaking, a cubic spline with T knots used a total of $4 + T$ degrees of freedom (Gareth et al, 2017), and a degree$-d$ spline is generally a piecewise degree$-d$ polynomial, with continuity in derivatives up to degree $d - 1$ at each knot.

Since each of the Figures 1 to 3 has a single knot at age = 50, based on remark V therefore, more knots could be added to enforce continuity at each knot.

Natural Splines

Natural spline is a restricted cubic spline that models a non-linear relationship with piecewise cubic polynomials with additional boundary constraints (Roksolana, 2019).

One setback of cubic splines is having high variance when the value of X is either very small or very large outside the range of the predictor which the natural spline overcame by enforcing linearity at the boundary (in the region where X is smaller than the smallest knot, or larger than the largest knot). This implies that, natural splines broadly yield more stable estimates at the boundaries. Figure 4 depicts the cubic spline's fit (solid blue curve) to the Wage data with three knots (Gareth *et al*., 2017). Observe that the confidence bands (the dotted blue curve) in the boundary region appear fairly wild.

Figure 4 also displays natural cubic spline as a red curve. Observe that the corresponding confidence intervals (the dotted red curve) are now narrower.

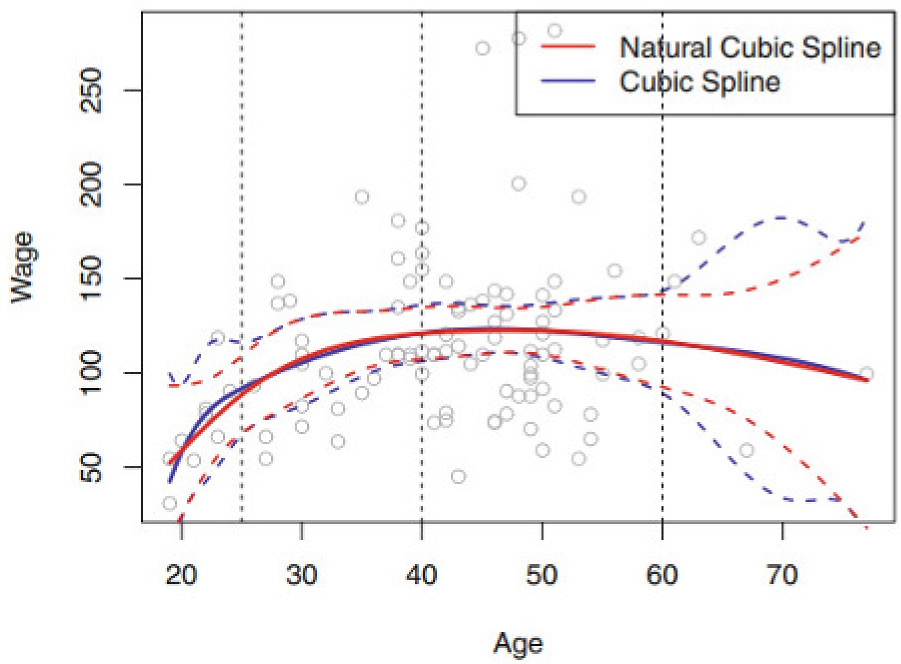

Fig. 4: A cubic spline and a natural cubic spline, with three knots, fit to a subset of the Wage data

Modeling and Representing Regression Splines

According to Gareth et al (2017), the basis model generally can be used to represent a regression spline. Taking an instance, a cubic spline with T knots can be modeled as

$$y_i = m_0 + m_1 b_1(x_i) + m_2 b_2(x_i) + m_3 b_3(x_i) + \ldots + m_{t+3} b_{t+3}(x_i) + \epsilon_i \qquad (2)$$

for an appropriate choice of basis functions $b_1, b_2, \ldots, b_{t+3}$. The model in equ. 2 can then be fit using least squares.

Representing a cubic spline in equ 2, a basis for a cubic polynomial—namely, x, x^2, x^3 is started off with, then one truncated power basis function per knot is added. A truncated power basis function is defined as

$$h(x, \xi) = (x - \xi)_+^3 = \begin{cases} (x - \xi)^3 & if\ x > \xi \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where ξ is the knot.

Remark VI: Adding a term of the form $m_4 h(x, \xi)$ to the model in equ. 1 for a cubic polynomial will lead to a discontinuity in only the third derivative at ξ; the function will remain continuous, with continuous first and second derivatives, at each of the knots.

This implies that, to fit a cubic spline with K knots, we perform least squares regression with an intercept and $3 + T$ predictors, of the form

$X, X^2, X^3, h(X, \xi_1), h(X, \xi_2),$
$. . . , h(X, \xi_t)$, where $\xi_1, . . . , \xi_t$ are the knots. This results to estimating a total of $T + 4$ regression coefficients. Hence, fitting a cubic spline with T knots used $T + 4$ degrees of freedom.

Motivations

According to John (2003), a piecewise polynomial improved polynomials regression by constraining its segments to at least have the same value at the segment boundaries. And the works of (Yi et al, 2002) supported adopting piecewise polynomial in interpolating adjacent time intervals for high prediction accuracy. In time sensitive applications and embedded systems, approximation of curvilinear profiles extraction from 3D point clouds, and realistic petroleum production optimization, piecewise polynomial facilitated optimization problems (Song and Lee, 2007), (Chiara *et al.*, 2023), (Bjarne and Brage. 2020).

Cubic spline regression has the potential of eliminating curves that exceed the limit where the error is large enough to fit the data characteristics. And the best cubic spline nonparametric regression model is influenced by the optimal knot point selection which is based on the minimum generalized cross validation (GCV) value (Mutia et al, 2022).

A landmark based approach, using five different interpolating polynomials (linear, cubic convolution, cubic spline, PCHIP, and Makima) for modeling of lung field region in 2D chest X-ray images revealed that the cubic spline method is of at least seven intermediate semilandmark points which does not result in better performance in terms of accuracy or execution time (Rohit *et al.*, 2021).

CONCLUSION

Considering the three regression splines' basic functions that extended the polynomial regression and step-wise piece function, the study revealed that the coefficients of a piecewise cubic polynomial differ in different parts of the range of X at the knots. Hence, it is not advised to use piecewise cubic polynomial because having more knots will lead to more flexible piecewise polynomial causing discontinuous fitted curves and higher degree of freedom. The study also showed that the limitations of the piecewise cubic polynomial were resolved by using cubic spine where the fitted curves were constrained using first and second derivatives, making the cubic spline to be continuous with the reduction in the degree of freedom.

REFERENCES

Bjarne, G. and Brage, R. K. (2020). Mathematical programming formulations for piecewise polynomial functions. *Journal of Global Optimization*, 77: 455–486.

Brendan, A. (2022). Cubic Splines: The Ultimate Regression Model. Towards Data Science. https://towardsdatascience.com/cubic-splines-the-ultimate-regression-model-bd51a9cf396d. Accessed 28/08/2024

Carl, D. B. (2002). Piecewise Polynomials. Handbook of Computer Aided Geometric Design. https://www.sciencedirect.com/topics/computer-science/piecewise-polynomial. Accessed 29/10/2024

Chiara, R., Ulderico, F., Bianca, F. and Silvia, B. (2023). Piecewise polynomial approximation of spatial curvilinear profiles using the Hough transform. *Applied Mathematics and Computation*, 457: https://doi.org/10.1016/j.amc.2023.128213

Dare, W. M. and Adeleke, K. A. (2021). Spline Regression Analysis with Application to Clinical Data: A Review. Nigerian Journal of Scientific Research. 20(5) 659-668.

Elhakeem A., Hughes, R.A., Tilling, K., Cousminer, D. L., Jackowski, S. A., Cole, T. J., Kwong, A. S. F., Li, Z., Grant, S. F. A., Baxter-Jones, A. D. G., Zemel, B. S., Lawlor, D. A. (2022). Using Linear and Natural Cubic Splines, SITAR, and Latent Trajectory Models to Characterise Nonlinear Longitudinal Growth Trajectories in Cohort Studies. *BMC Med. Res. Methodol*., 22(1): 68. https://doi.org/10.1186/s12874-022-01542-8

Gareth, J., Daniela, W., Trevor, H. and Robert, T. (2017). An Introduction to Statistical Learning with Application in R. Springer, New York. 434pp.

Hernandez, H. (2022). Cubic Spline Regression using OAT Optimization. *Fors. Chem. Research Reports*, 7(2022-13): 1 - 34.

John, N. S. (2003). V.I.B Piecewise Polynomials and Splines. Encyclopedia of Physical Science and Technology. https://www.sciencedirect.com/referencework/9780122274107/encyclopedia-of-physical-science-and-technology

Mutia, M., Dewi, R. S. S. and Putranto, H. U. (2022). Application of the cubic spline nonparametric regression model to human development index data in Indonesia. *AIP Conference Proceedings, 2479 020030*. https://doi.org/10.1063/5.0099712

Ogoke, P., Nduka, C. and Soyinka, A. (2016). Cubic Spline Regression: An Application to Early Bipolar Disorder Dynamics. Open Journal of Statistics. 6 1003-1009. doi: 10.4236/ojs.2016.66080.

Rohit, K., Bhattacharya, S. and Govind, M. (2021). Exploring Optimality of Piecewise Polynomial Interpolation Functions for Lung Field Modeling in 2D Chest X-Ray Images. Sec. *Optics and Photonics*, 9: https://doi.org/10.3389/fphy.2021.770752

Roksolana, P. (2019). Using Natural Splines in Linear Modeling. https://www.clayford.net/statistics/using-natural-splines-in-linear-modeling/. Accessed 28/08/2024

Song, J. and Lee, B. (2024). Optimal Piecewise Polynomial Approximation for Minimum Computing Cost by Using Constrained Least Squares. Sensors. 24 3991. https://doi.org/10.3390/s24123991

Yi, W., Youpeng, Y. and Jianjun, J. (2002). New predictor-corrector methods based on piecewise polynomial interpolation for milling stability prediction. *Machining Science and Technology*, 24(5): 688-718. https://doi.org/10.1080/10910344.2020.1752235