

Cubic Spline Regression: An Application to Early Bipolar Disorder Dynamics

Petronilla Uchenna Ogoke^{1*}, Chinaka Ethelbert Nduka¹, Ajibola Taiwo Soyinka²

¹Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Nigeria

²Department of Research and Training, Federal Neuro-Psychiatric Hospital Aro, Abeokuta, Nigeria

Email: *uchedubem@yahoo.com, etelnduka@yahoo.com, soyinkataiwo@yahoo.ie

How to cite this paper: Ogoke, P.U., Nduka, C.E. and Soyinka, A.T. (2016) Cubic Spline Regression: An Application to Early Bipolar Disorder Dynamics. *Open Journal of Statistics*, 6, 1003-1009.

<http://dx.doi.org/10.4236/ojs.2016.66080>

Received: May 17, 2016

Accepted: November 8, 2016

Published: November 14, 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Owing to the fact that the major challenge of predicting the risk of having bipolar is the absence of a gold standard to distinguish between true cases and false positive; this study employed the extension of cubic spline function to the multinomial model to explore the risk tendency of unnoticed early bipolar across three different groups of mood disorder. The intermediate group was used to accommodate for false negative and false positive while mapping the true value of bipolar risk tendency across the three groups to a scale. Hence for all distributions of “yes” ticked in a mood disorder questionnaire, the study predicts the bipolar risk tendency while simultaneously accommodating for the patients response bias. The coefficients of the polynomial are obtained using the maximum likelihood method. The spline graph reveals 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.

Keywords

Bipolar, Gold Standard, Multinomial Model, Response Bias, Risk Tendency

1. Introduction

Spline is a numeric function that is piecewise-defined by polynomial functions and which possesses a high degree of smoothness at the places where the polynomial pieces connect (which are known as knots) Chen 2009 [1], Judd 1998 [2]. As noted by Harrell *et al.* (1988) [3], splines are smooth functions that can assume virtually any shape, and the most useful type of spline is generally a cubic spline function, which is restricted to be smooth at the junction of each cubic polynomial. A restricted cubic spline model has been used in epidemiological studies and it is often applied to nonlinear dose-response data as noted by Larsson and Orsini (2011) [4]. Takahashi *et al.* (2013) [5] applied a re-

stricted cubic spline with three knots recently to a potential nonlinear association that was depicted as a J-shaped curve based on the likelihood-based assignment of values to grouped intervals of exposure. The most commonly used spline is the cubic spline functions. Spline regression has been widely used in different spheres.

Berberoglu and Berberoglu (2011) [6] applied cubic spline regression to model the structural shifts in exchange rate in Turkey from 1987-2008. Cubic spline regression was used to expose structural changes which resulted because of the economic policies. They built different cubic spline regression and picked the most significant of the models. In their study, cubic spline models was also identified as the most powerful and important weapon on the existence of structural shifts or changes in time series. They also pointed out how predicted sum of squares statistics residual can improve the analysis in cubic spline models. The usefulness of spline regression cannot be over emphasized because it represents a less biased and more efficient alternative to standard linear, curvilinear, or categorical analyses of continuous exposures and confounders in observational study [7]. Benefits of restricted cubic and quadratic splines have been described in the epidemiological and biomedical literature ([3] and [8]).

In order to draw our attention to some of the challenging issues to health, this paper thus addresses the two major problems of the MDQ which are

- The problem of which rule is the best to decide bipolar risk tendency.
- The problem of response bias from patients most especially when the patient's statement is incoherent or contradicts that of the care giver.

The paper use the cubic spline curve to explore the bipolar disorder dynamics in relation to the behaviour of MDQ scores across the false negative and false positive knots and also estimate the real MDQ score for patient suspected of incoherent statements.

The importance of Mood Disorder Questionnaire (MDQ) developed by a committee of mental health experts in predicting the risk of bipolar disorder has been a course of concern for psychiatrist and researchers in mental health. The present rule of deciding bipolar risk is based on patients' response during clerking; meeting some diagnostic criteria (Hirschfeld *et al.* 2000 [9]). This is obviously subject to response bias as many patients may resort to lying just to avoid admission or avoid been stigmatized.

2. Methodology

Let the function

$$f(x, y, z) = \frac{n!}{x!y!z!} \beta_1^x \beta_2^y \beta_3^z \quad (1)$$

be the trinomial distribution (a special case of the multinomial distribution); where x and y are non-negative integers with $x + y \leq n$ and $z = n - x - y$. Also β_1, β_2 and β_3 are the positive proper fraction with constrains $\beta_1 + \beta_2 + \beta_3 = 1$ and let $f(x, y) = 0$ elsewhere (Hogg and Craig [10]), (Mood *et al.* [11]).

The natural log-likelihood function of (1) for k th trials is given as

$$\log L(x_i, y_i) = \sum_{i=1}^k \log \left[\frac{n!}{x_i!y_i!z_i!} \right] + \sum_{i=1}^k x_i \log \beta_1 + \sum_{i=1}^k y_i \log \beta_2 + \left[n - \sum_{i=1}^k (x_i + y_i) \right] \log \beta_3 \quad (2)$$

Note that n is fixed and z_i is a random variable given as $z_i = n - x_i - y_i$ which in reality is a partition beyond the space of y_i .

Expressing (2) as a family of exponential class joint probability density function, then we have

$$L(x_i, y_i) = \exp \left[\frac{k \log(n!) - \sum_{i=1}^k \log(x_i! y_i! z_i!) + \sum_{i=1}^k x_i \log \beta_1 + \sum_{i=1}^k y_i \log \beta_2 + \left[n - \sum_{i=1}^k (x_i + y_i) \right] \log \beta_3}{\phi} \right] \quad (3)$$

Comparing (3) to the general form of the exponential class

$f(u_i | \beta, \phi) = \exp \left[\frac{P(\beta^T) R(u_i) - Q(\beta)}{\phi} + S(u_i, \phi) \right]$ we observed that the natural parameter $P(\beta^T)$ is the vector of the model parameters $[\log \beta_1, \log \beta_2, \log \beta_3]^T$; the sufficient statistics $R(u_i)$ is the vector of the model matrix $\left[\sum_{i=1}^k x_i, \sum_{i=1}^k y_i, n - \sum_{i=1}^k (x_i + y_i) \right]$; the base measure is $S(u_i, \phi) = k \log(n!) - \sum_{i=1}^k \log(x_i! y_i! z_i!)$, the log-partition function is $Q(\beta) = 0$ while the scale parameter is $\phi = 1$ (Hogg and Craig [9], p. 231).

Hence the monotone and differentiable link functional relationship (g) between the expected response value of the random component $L(x_i, y_i)$ and the systematic (linear predictor) component is

$$E[\log \{L(x_i, y_i)\}] = g^{-1} \left[\sum_{i=1}^k x_i \log \beta_1 + \sum_{i=1}^k y_i \log \beta_2 + \sum_{i=1}^k (n - x_i - y_i) \log \beta_3 \right] \quad (4)$$

Supposing $0 < x_i \leq t_1$, $t_1 < y_i \leq t_2$ and $z_i > t_2$ are partitions according to the earlier assumptions of the committee of mental health experts (Hirschfeld *et al.* 2000 [9]); then (4) is a piece-wise linear mixture of probability density for k th trials where each trial is partitioned into three groups of distinct intervals.

So extending the linear function representation in (4) to its cubic polynomial form, we have

$$E[\log \{L(x_i, y_i)\}] = g^{-1} \left(\begin{bmatrix} x_1^3 & x_1^2 & x_1^1 & 1 \\ x_2^3 & x_2^2 & x_2^1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_k^3 & x_k^2 & x_k^1 & 1 \end{bmatrix} \begin{bmatrix} \log \beta_{13} \\ \log \beta_{12} \\ \log \beta_{11} \\ \log \beta_{10} \end{bmatrix} + \begin{bmatrix} y_1^3 & y_1^2 & y_1^1 & 1 \\ y_2^3 & y_2^2 & y_2^1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_k^3 & y_k^2 & y_k^1 & 1 \end{bmatrix} \begin{bmatrix} \log \beta_{23} \\ \log \beta_{22} \\ \log \beta_{21} \\ \log \beta_{20} \end{bmatrix} + \begin{bmatrix} z_1^3 & z_1^2 & z_1^1 & 1 \\ z_2^3 & z_2^2 & z_2^1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ z_k^3 & z_k^2 & z_k^1 & 1 \end{bmatrix} \begin{bmatrix} \log \beta_{33} \\ \log \beta_{32} \\ \log \beta_{31} \\ \log \beta_{30} \end{bmatrix} \right) \quad (5).$$

Hence Equation (5) is the cubic polynomial spline equation for the k th trials trinomial model over the boundary conditions

$$\begin{cases} 0 < x_i \leq t_1 \\ t_1 < y_i \leq t_2 \\ t_2 < z_i \end{cases} \quad (6)$$

To estimate the parameters, we maximized (5) by differentiating piece-wisely with respect to the parameters. The resulting homogeneous matrix $(A - \lambda)v = 0$ is then solved via characteristic function $|A - \lambda I|$ for the eigen-values λ and linear function $Av = \lambda v$ for the eigen vectors v (See Johnson and Wichern (2007) [12]). A is the square of the rectangular matrices in (5).

3. Application

A total of seven hundred students that filled the mood disorder questionnaire and agreed to have landed in at least minor problems as a result of their irrational behaviour over a period of three months were used in the study. None of the respondents has reported in any psychiatric facility before. The observed scores were then simulated to a sample size of $k = 1416$ each for the three different category of bipolar risk. Note that the scores are discrete. That is the number of yes score in a single trial (a single mood disorder questionnaire filled) is partitioned into three groups which are Bipolar NOS (Bipolar Not Otherwise Specified) $1 < x_i \leq 3$, Bipolar I $3 < y_i \leq 7$, and Bipolar II $z_i > 7$. The maximum number of yes scores in a single trial is fixed $n = 13$ (Hirschfeld *et al.* 2000 [9]).

4. Result

Cubic spline equation for each of the mood disorder grouping.

Mood Disorder Grouping	Highest Eigen-value (%)	Associated Eigen-vectors	Natural Logarithm β
Bipolar NOS $1 < x_i \leq 3$	$3.97847 \exp(5)$ 99.4%	$-0.04435436; -0.5047424 \ 0.74679367; 0.43077490$	$0.9566; 0.60366 \ 2.110223; 1.538449$
Bipolar I $3 < y_i \leq 7$	$1.902910 \exp(6)$ 99.74%	$-0.01976824; -0.3427822 \ 0.7881406; 0.51082675$	$0.9804259; 0.70979 \ 2.1993; 1.66667$
Bipolar II $z_i > 7$	$1.42 \exp(7)$ 99.9%	$-0.00643049; -0.1795336 \ 0.8779754; 0.44371793$	$0.9935901; 0.83566 \ 2.4060; 1.5585$

Bipolar NOS: Bipolar Not Otherwise Specified.

5. Discussion of Result

Figure 1, reveals two maximum points (Knot 3 and Knot 7) and two minimum points (Knot 4 and Knot 8) indicating three separate density groups. The intermediate group is at knot 4 and knot 8, with false negative interval between knot 4 and knot 5 and false positive interval between knot 6 and knot 7. Any patient with a score of knot 3 belongs to the Bipolar NOS group. The Bipolar II group begins at knot (4). So there is uncertainty of classification between knot 3 and knot 4 and likewise between knot 7 and knot 8. So Bipolar I begin from Knot 8.

Figure 2 is useful to remove the problem of patients' bias. Any incoherent patient with a claim of MDQ score 2 can be adjusted for bias via the graph in **Figure 2**. From the graph, the patient's real estimated score in Bipolar II group is between 5 and is tending towards 8; while its score in the Bipolar I group is within the interval 8 - 9. This will guide the psychiatrist on the best therapy to give to the patient despite his/her incoherent claim.

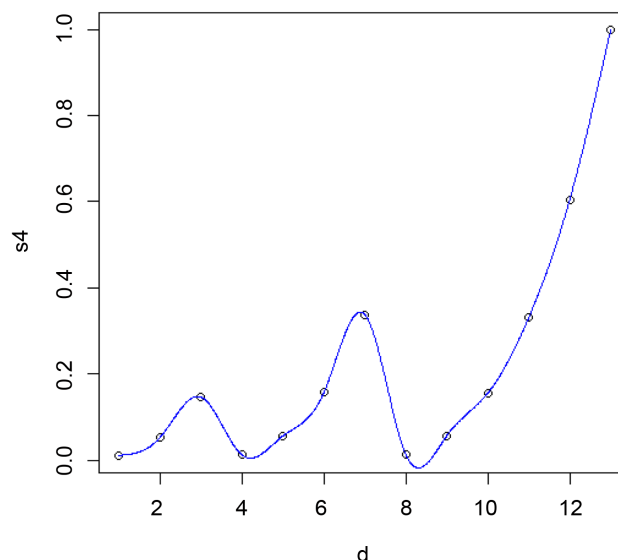


Figure 1. The general bipolar mood disorder dynamics.

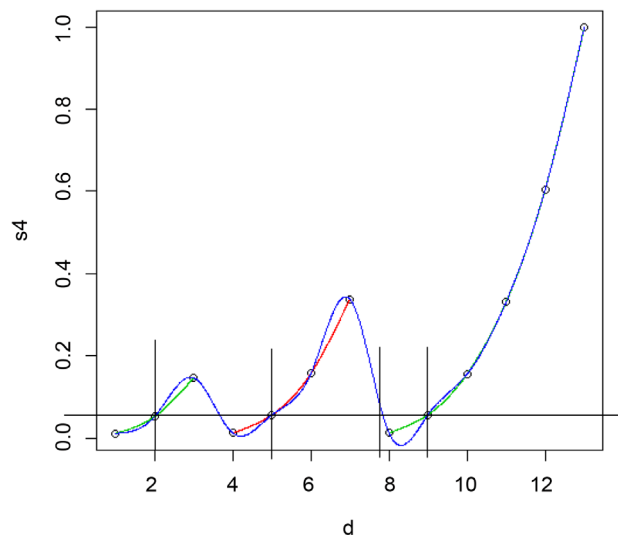


Figure 2. Combined curve of the general and the individual bipolar mood disorder dynamics.

Figure 3 is the plot of continuous scores for the MDQ as against the initial assumption of discrete scores as agreed by the experts that developed the questionnaire. The plot revealed that Bipolar NOS group is truly before the knot 4. While knot 4 marks the beginning of classifying bipolar patient to the Bipolar II group. However, the fluctuation in the bipolar disorder risk tendency is well captured in **Figure 3** as against what we have in **Figure 2**. The bipolar disorder relative risk tendency is approximately 0.22 at knot 5, this dropped to 0.158 at knot 6 (a drop of about 28.2%). This implies that despite the increment in the mood disorder score, the bipolar disorder relative risk tendency is not directly increasing. Also, a more pronounced drop occurs between knot 9 and knot 10 (0.378 to 0.08 respectively representing a drop of about 78.8%).

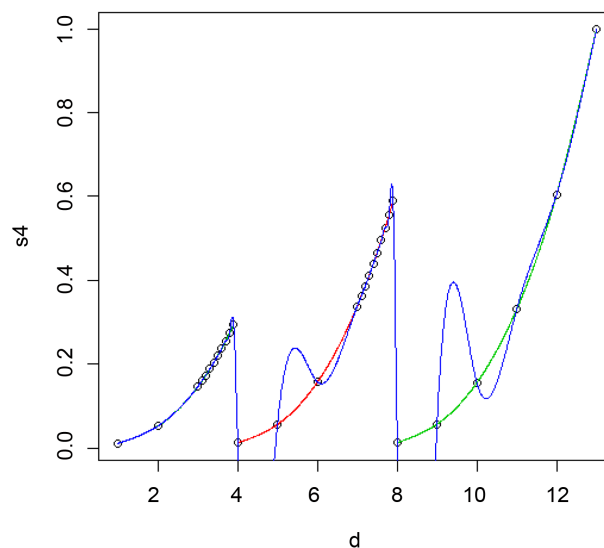


Figure 3. Bipolar disorder dynamics assuming continuous score for MDQ.

6. Conclusion and Recommendation

The fluctuation in the dynamics of bipolar disorder at different knots is responsible for the unnoticed prolonged build up of bipolar disorder in the body. So this study recommends that any patient that has a MDQ score of at least 4 is a potential bipolar patient and should be treated accordingly given the necessary therapy. This is because bipolar disorder risk tendency begins to build up unnoticed from knot 4. Anyone with MDQ score of three below should be monitored appropriately until the level of mood disorder can be specified. We also recommend **Figure 2** in addressing incoherent claim in patients' submission. Finally we recommend proper awareness of mood control among individuals without psychiatric history majorly within students in tertiary institution.

References

- [1] Chen, W. (2009) Feedback, Nonlinear, and Distributed Circuits. CRC Press, Book News, Inc., Portland, OR, 9-20.
- [2] Judd, K.L. (1998) Numerical Methods in Economics. MIT Press, Cambridge, US, 225.
- [3] Harrell, F.E., Lee, K.L. and Pollock, B.G. (1998) Regression Models in Clinical Studies Determining Relationships between Predictors and Response. *Journal of National Cancer Institute*, **80**, 1198-1202. <http://dx.doi.org/10.1093/jnci/80.15.1198>
- [4] Larsson, S.C. and Orsini, N. (2011) Coffee Consumption and Risk of Stroke: A Dose-Response Meta-Analysis of Prospective Studies. *Amsterdam Journal of Epidemiology*, **174**, 993-1001. <http://dx.doi.org/10.1093/aje/kwr226>
- [5] Takahashi, K., Nakao, H. and Hattori, S. (2013) Cubic Spline Regression of J-Shaped Dose-Response Curves with Likelihood-based Assignments of Grouped Exposure Levels. *Journal of Biometrics and Biostatistics*, **4**, 181. <http://dx.doi.org/10.4172/2155-6180.1000181>
- [6] Berberoglu, B. and Berberoglu, C.N. (2011) Modeling the Structural Shifts in Real Exchange Rate with Cubic Spline Regression. *International. Journal of Business and Social Science*, **2**,

60. www.ijbssnet.com
- [7] Wu, W. (2009) An Application of Spline Regression to Dose-Response Analysis in Observational Study. Cancer Biostatistics Center, Preston Building Nashville US.
- [8] Greenland, S. (1995) Dose-Response and Trend Analysis in Epidemiology: Alternatives to Categorical Analysis. *Epidemiology*, **6**, 356-65.
<http://dx.doi.org/10.1097/00001648-199507000-00005>
- [9] Hirschfeld, R.M., Williams, J.B., Spitzer, R.L., Calabrese, J.R., Flynn, L., Keck Jr., P.E., Lewis, L., McElroy, S.L., Post, R.M., Rapport, D.J., Russell, J.M., Sachs, G.S. and Zajecka, J. (2000) Development and Validation of a Screening Instrument for Bipolar Spectrum Disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry*, **157**, 1873-1875.
- [10] Hogg, R. and Craig, A. (1970) Introduction to Mathematical Statistics. 3rd Edition, 91, 231.
- [11] Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) Introduction to the Theory of Statistics. McGraw-Hill Series in Probability and Statistics. Kinsport Press, Inc., USA.
- [12] Johnson, R.A. and Wichern, D.W. (2007) Applied Multivariate Statistical Analysis. Prentice-Hall, Inc., Englewood Cliffs.

Appendix (R 3.22)

Appendix: THE PROGRAM TO CONSTRUCT THE SPLINE CURVE

```
> a<-c(1,2,3)
> b<-c(4,5,6,7)
> c<-c(8,9,10,11,12,13)
> h<-array(a, dim=c(1,1416))
> I <-array(b, dim=c(1,1416))
> j<-array(c, dim=c(1,1416))
> s1<-0.9566+0.60366*h+2.110223*(h^2)+1.5384492*(h^3)
> s2<-0.9804259+0.709728*(i-3)+2.199303*((i-3)^2)+1.66667*((i-3)^3)
> s3<-0.9935901+0.8356599*(j-7)+2.4060235*((j-7)^2)+1.5584908*((j-7)^3)
> d<-c(h,i,j)
> s4<-c(s1,s2,s3)
> plot(d,s4)
> g <- plot(d,s4)
> lines(spline(h, s1, n = 1416, method = "natural"), col = 3)
> lines(spline(i, s2, n = 1416, method = "natural"), col = 2)
> lines(spline(j, s3, n = 1416, method = "natural"), col = 3)
> lines(spline(d, s4, n = 1416, method = "natural"), col = 4)
```



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ojs@scirp.org