# Deriving Penalized Splines For Estimation Of Time Varying Effects In Survival Data

Albert Otieno Orwa, George Otieno Orwa, Romanus Odhiambo

**Abstract:** The major interests of survival analysis are either to compare the failure time distribution function or to assess the effects of covariate on survival via appropriate hazards regression models. Cox's proportional hazards model (Cox, 1972) is the most widely used framework, the model assumes that the effect on the hazard function of a particular factor of interest remains unchanged throughout the observation period (Proportionality assumption). For a continuous prognostic factor the model further assumes linear effect on the log hazard function (Linearity assumption). Assumptions that many authors have found to be questionable when violated since they may result to biased results and conclusions and as such non-linear risk functions have been suggested as the suitable models.In this paper, we propose a flexible method that models dynamic effects in survival data within the Cox regression framework. The method is based on penalized splines. The model offers the chance to easily verify the presenceof PH and time-variation. We provide a detailed analysis and derivation of the penalized splines in the context of survival data.

**Key Words:**Non-linear, Penalized splines, Proportional Hazard, Survival analysis

## 1.0 Introduction

Survival analysis entails a wide variety of methods aimed at analyzing the timing of events. Many researchers are able to model and assess why certain subjects are exposed to a higher risk of experiencing an event of interest such as death, development of an adverse reaction or relapse of a given disease (e.g. Cancer). Cox proportional hazard(CPH) model is the most popular regression model used for the analysis of survival data. The model allows testing for the differences in survival times of more than one group of interest and compares the cumulative probability of the events, while adjusting other influential covariates. It is a semi-parametric model that makes fewer assumptions than a typical parametric method. One of the assumptions of the Cox model is the linearity of the covariates variables on the log hazard function. The non-flexibility of these methods subjects the model to biasness. For instance, they assume independence of covariates that affect the hazard rate. The CPH model also makes assumptions that the factors are linear yet findings have indicated that some prognostic factors (for example, body mass index) have non-linear effect on breast cancer survival and/or prognosis (Gray, 1994). Based on this, cox proportional model poses a problem in analyzing time-to-event data;

i) It is complex to relate the variables to the outcome.
ii) The variables interact with each other.
iii) It is not possible to apply the assumption of proportionality of the hazards to the data.


- *Albert Otieno Orwa1, George Otieno Orwa2, and Romanus Odhiambo3*
- *Pan African University Institute of Basic Science, Technology and Innovations P.O Box 62000 00200 Nairobi-Kenya*
- *Department of Statistics and Actuarial Sciences Jomo Kenyatta University of Agriculture and Technology*
- *P.O Box 62000-00200, Nairobi-Kenya*


This could possibly lead to biased risk estimates and as such distorting the findings. (Hastie, T & Tibshirani, R, 1990) have shown that a better choice is to use smoothing splines, where knot selection is automatic based on a mean squared criterion. With smoothing splines, the user only need to select the level of smoothness, which is done by selecting the degrees of freedom for each spline fit. Time-varying effects (TVEs) of prognostic factors have been detected in a variety of medical fields. Gore et.al (1984) presents a classical example discussing this issue for several covariates that relate to breast cancer. In the same disease, the effects of oestrogen receptor and tumour size have been reported to change over time (Hilsenbeck, S. G., et.al, 1998). Other examples have been established to include the effects of prothrombin time in primary biliary cirrhosis (Abrahamowicz, M., , T. MacKenzie,, & J. M. Esdaile, 1996), the Karnofsky performance status in ovarian cancer studies (Verweij & Houwelingen, 1995) and diabetes on mortality after coronary artery bypass graft surgery (Gao C, , Yang M,, Wu Y, , & et al, 2006).

## 2.0 Review of regression models

### 2.1 Regression model

A regression model is basically of the form;

$$Y = f(X_1, X_2, \dots, X_n) + \varepsilon \qquad (1)$$

Where, Y is the response (dependent) variable $X_1, X_2, \dots, X_n$ are the predictor (independent) variables, $\varepsilon$ is the error or simply the difference between the model and the actual values. The regression model aims at minimizing the error ($\varepsilon$) for all the values of $Y$ without introducing extraneous and arbitrary random variables. For a single predictor variable (univariate variable) we have;
$Y = f(X) + \varepsilon$ for some function f.

### 2.2 Simple linear regression

Simple linear regression or ordinary least squares (OLS) fits a straight line to the dataset of interest. It is given as;

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2)$$

Where $\varepsilon$ is the error term (accounts for difference between the observed and predictedvaluesof $Y$). We make assumptions that the error has a mean zeroand a constant variance $\sigma^2$, and isidentically independently distributed (iid). By this we meanthat each error term is centered about the line of best fit (with mean zero) and that there is a constant amount of deviation of the error terms from the line of best fit (with constant variance).To find the line of best fit through the scatter plot of $(X, Y)$ values, we actually aimat minimizing the error term for all the values of $y$. We then modify our equation as follows;

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i (3)$$

The fitted or predicted value thus becomes;

$$\hat{y}_i = \beta_0 + \beta_1 x_i \quad (4)$$

This simply implies that; $y_i = \hat{y}_i + \varepsilon_i$

We can rewrite the model by solving the error term as;

$$\varepsilon_i = y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_i \quad (5)$$

Selecting the values of $\beta_0$ and $\beta_1$ that minimizes the total error as much as possible. We have;

$$\text{Min} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \quad (6)$$

Taking the partial derivatives and set them to zero

$$\frac{\partial}{\partial \beta_0} = \sum_{i=1}^{n} -2(y_i - \beta_0 - \beta_1 x_i) \quad (7)$$

$$\frac{\partial}{\partial \beta_1} = \sum_{i=1}^{n} -2x_i(y_i - \beta_0 - \beta_1 x_i) \quad (8)$$

The above equations yield the following two normal equations;

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i \quad (9)$$

$$\sum_{i=1}^{n} x_i y_i = \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} {x_i}^2 \quad (10)$$

These are two equations in two unknowns. We can thus solve for $\beta_0$ and $\beta_1$ yielding;

$$\beta_0 = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i\right) = \bar{y} - \beta_1 \bar{x} \quad (11)$$

$$\beta_1 = \frac{\sum_{i=1}^{n} y_i - n\beta_0}{\sum_{i=1}^{n} x_i} \quad (12)$$

### 2.3 Derivation of spline regression model

Spline regression is a regression model with piecewise continuous polynomial function. We intend to derive penalized spline. Considering a simple linear model (4), and applying the concept of algebra we have; $\hat{y} = X\beta$ which can be rewritten in matrix form as follows;

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \; X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \beta = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \text{ with}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (13)$$

Clearly $\hat{y}$ is a unique linear combination of the x-values and 1, the basis is thus x and 1.

### 2.4 Penalized Splines

Using penalization criteria we choose Q such that;

$$\sum_{i=1}^{k} b_i^2 < Q \quad (14)$$

The above equation represents a minimization criterion since it reduces the overall effect of individual piecewise functions and avoids over-fitting the data. We can formally state the minimization criterion as minimizing the equation given below;

$|y - X\beta|^2$subject to $\beta^T D\beta \leq Q$, where;

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 0_{2\times2} & 0_{2\times k} \\ 0_{k\times2} & I_{k\times k} \end{bmatrix} (15)$$

Applying Lagrange Multiplier results an equation which is equivalent to minimizing;

$$|y - X\beta|^2 + \lambda^2 \beta^T D\beta \quad (16)$$

$$\text{for some } \lambda \geq 0 \qquad \text{w.r.t } \beta$$

We now aim at solving the optimal $\hat{\beta}$ for any given value of $\lambda$. We need to derive two common matrix equations and show that;

i) $\frac{\partial(a^T\beta)}{\partial\beta} = a \quad (17)$

ii) $\frac{\partial(\beta^T A\beta)}{\partial\beta} = 2A\beta \quad (18)$

Where $a$ is a $2 \times 1$ vector, A is a $2 \times 2$ symmetric matrix, $\beta = [\beta_0 \beta_1]^T$, and the partial $g(\beta)$ w.r.t $\beta$ is;

$$\frac{\partial g(\beta)}{\partial \beta} = \begin{bmatrix} \partial g(\beta)/\partial\beta_0 \\ \partial g(\beta)/\partial\beta_1 \end{bmatrix}$$

i) By multiplication we know that;

$$a^T\beta = a_1\beta_0 + a_2\beta_1$$

$$\therefore \frac{\partial(a^T\beta)}{\partial\beta} = \begin{bmatrix} \partial(a_1\beta_0 + a_2\beta_1)/\beta_0 \\ \partial(a_1\beta_0 + a_2\beta_1)/\beta_1 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = a$$

ii) By multiplication we also know that;

$$A\beta = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} a_1\beta_0 & a_2\beta_1 \\ a_2\beta_0 & a_3\beta_1 \end{bmatrix}$$
$$\beta^T A\beta = a_1\beta_0^2 + 2a_2\beta_0\beta_1 + a_3\beta_1^2$$

Using partial derivatives we obtain;

$$\frac{\partial(\beta^T A\beta)}{\partial\beta} = \begin{bmatrix} \dfrac{\partial(a_1\beta_0^2 + 2a_2\beta_0\beta_1 + a_3\beta_1^2)}{\beta_0} \\ \dfrac{\partial(a_1\beta_0^2 + 2a_2\beta_0\beta_1 + a_3\beta_1^2)}{\beta_1} \end{bmatrix}$$

$$= \begin{bmatrix} 2a_1\beta_0 + 2a_2\beta_1 \\ 2a_2\beta_0 + 2a_3\beta_1 \end{bmatrix}$$

$$= 2\begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$
$$= 2A\beta$$

**2.5Deriving the Penalized Spline Solution**

The solution to the penalized spline will involve minimizing (16), that is solving when all the partial derivatives with respect to $\beta_0$ and $\beta_1$ are zero. This could be represented mathematically as;

$$\frac{\partial}{\partial\hat{\beta}}\left(\|y - X\hat{\beta}\|^2\right) + \frac{\partial}{\partial\hat{\beta}}\left(\lambda^2\hat{\beta}^T D\hat{\beta}\right) = 0 \qquad (19)$$

Since differentiation is linear, we are able to split (19) into two parts With the two identities already proved we get;

$\frac{\partial}{\partial\hat{\beta}}\left(\|y - X\hat{\beta}\|^2\right) = 2X^T(y - X\hat{\beta})$where$X^T y$ is the vector $a^T$ and $X^T X$ is the matrix A. We also have;

$\frac{\partial}{\partial\hat{\beta}}\left(\lambda^2\hat{\beta}^T D\hat{\beta}\right)$which by linearity of differentiation, $\lambda$ gets factored out leaving;

$\lambda^2\frac{\partial}{\partial\hat{\beta}}\left(\hat{\beta}^T D\hat{\beta}\right)$with D being symmetrical, we gain apply the differentiation identities to get;

$\frac{\partial}{\partial\hat{\beta}}\left(\lambda^2\hat{\beta}^T D\hat{\beta}\right) = 2\lambda^2 D\hat{\beta}$, we finally combine the partial derivatives to get;

$$\lambda^2 D\hat{\beta} - X^T\left(y - X\hat{\beta}\right) = 0 \qquad (20)$$

Clearly from linear algebra we can manipulate (20) to get;

$$\lambda^2 D\hat{\beta} = X^T\left(y - X\hat{\beta}\right)$$
$$\lambda^2 D\hat{\beta} = X^T y - X^T X\hat{\beta}$$
$$X^T X\hat{\beta} + \lambda^2 D\hat{\beta} = X^T y$$
$$\hat{\beta}(X^T X + \lambda^2 D) = X^T y$$
$$\hat{\beta} = (X^T X + \lambda^2 D)^{-1}X^T y$$

Now since we already have $\hat{\beta}$ and we know that $\hat{y} = X\hat{\beta}$ we now fit the penalized spline as follows;

$$\hat{y} = X(X^T X + \lambda^2 D)^{-1}X^T y \quad (21)$$

## 3.0 Smooth Hazard Model

### 3.1Fitting the Penalized spline (P-Spline)

Given the survival time $\tau_i$ for the ith observational unit, we define $C_i$ to represent the right censoring time; with $i = 1,2,\ldots,N$.We note that $Y_i = \min(\tau_i, C_i)$. We also define the censoring indicator, $\delta_i$ as follows;

$$\delta_i = \begin{cases} 1 & \text{if } \tau_i < C_i \\ 0 & \text{Otherwise} \end{cases} \qquad (22)$$

Now given a covariate $x_i$, which is independent of time and denoted by p − dimensional covariate vector for the ith observational unit, we can then model the hazard function as;

$$h(t, x_i) = \lambda_0(t)\exp\{x_i^T\beta_x(t)\} \qquad (23)$$

Where $\lambda_0$ is the baseline hazard, $\beta_x(t)$ is the vector of covariate effects that vary smoothly with survival time, t.The main idea is to estimate $\beta(t)$ smoothly by avoiding the tough parametric assumptions. A common approach to dealing with non-linear relationship is to approximate f by a polynomial of order m (Yuedong. W, 2011).For instance,

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_{m-1}x^{m-1} \quad (24)$$

Applying the Sobolev Space, $f \in W_2^m[a,b]$ we have

$$W_2^m[a,b] = \left\{ f: f, f', \ldots, f^{m-1} \text{ are absolutely continuous}, \int_a^b \left(f^{(m)}\right)^2 dx < \infty \right\} \qquad (25)$$

By Taylors theorem,

$$f(x) = \underbrace{\sum_{v=0}^{m-1}\frac{f^{(v)}(a)}{v!}(x-a)^v}_{\text{Polynomial of order } m} + \underbrace{\int_a^x \frac{(x-u)^{m-1}}{(m-1)!}f^{(m)}(u)du}_{\text{Rem }(x)} \; (26)$$

The polynomial regression in (24) ignores the remainder term Rem (x), it could be mere assumption that Rem (x) is negligible. The idea behind smoothing spline is simply to let data decide how large Rem (x) is going to be. Now using the least squares (LS) on $W_2^m[a,b]$, an infinite dimensional space, we have;

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 \qquad (27)$$

The distance measure between f and polynomial is,

$$\int_a^b \left(f^{(m)}\right)^2 dx \qquad (28)$$

We now estimate $f$ by minimizing $LS$ under the constraint say, $\rho$ which yields;

$$\int_a^b \left(f^{(m)}\right)^2 dx \le \rho \quad \text{where } \rho \text{ is some constant.}$$

We introduce a Lagrange multiplier in (27) and (28) so as to get Penalized Least Squares (PLS)

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i)^2 + \lambda\int_a^b \left(f^{(m)}\right)^2 dx \qquad (29)$$

$\int_a^b \left(f^{(m)}\right)^2 dx$is called the roughness penalty

If we consider now the Sobolev space $W_2^m[a,b]$ with a linear product

$$(f,g) = \sum_{v=0}^{m-1} f^{(v)}(a)g^{(v)}(a) + \int_a^b f^{(m)}\, g^{(m)}dx \ (30)$$

We further say that $W_2^m[a,b] = \mathcal{H}_0 \oplus \mathcal{H}_1$, where

$$\mathcal{H}_0 = \text{span}\,\{1, (x-a), \dots, (x-a)^{m-1}/(m-1)!\} \ (31)$$

$$\mathcal{H}_1 = \left\{ f\colon f^{(v)}(a) = 0, v = 0, \dots, m-1, \int_a^b \left(f^{(m)}\right)^2 dx < \infty \right\} (32)$$

Now (31) and (32) are RKHS's with the RKs

$$R_0(x,z) = \sum_{v=1}^{m} \frac{(x-a)^{v-1}(z-a)^{v-1}}{(v-1)!\,(v-1)!} \ (33)$$

$$R_1(x,z) = \int_a^b \frac{(x-u)_+^{m-1}(z-u)_+^{m-1}}{(m-1)!\,(m-1)!} du \quad (34)$$

$(x)_+$means that $\max(x,0)$

Looking at (31), it is clear that $\mathcal{H}_0$ contains a polynomial of order $m$ in the Taylor expansion. If we now denote $Q$ to be the orthonormal projection operator onto $\mathcal{H}_1$ and based on the definition of the inner product, the roughness penalty is; $\int_a^b \left(f^{(m)}\right)^2 dx = ||Qf||^2$which shows that $\int_a^b \left(f^{(m)}\right)^2 dx$ measures the distance between parametric polynomial space $\mathcal{H}_0$ and f. $\mathcal{H}_0$ has no penalized functions. The penalized least squares is thus;

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x))^2 + \lambda||Qf||^2$$

Where $\lambda$ is a smoothing parameter that controls the balance between the goodness-of-fit measured by the least squares and departure from the null space $\mathcal{H}_0$ measured by $||Qf||^2$. Functions in $\mathcal{H}_0$ are not penalized since

$||Qf||^2 = 0$when $f \in \mathcal{H}_0$.

## Conclusion

In the analysis of larger studies of censored data with long-term follow-up, the usual common standard techniques such as Cox model (Cox, 1972) may not be appropriate due to violation of the proportional hazard assumption that is caused by the time-varying effects. By ignoring the presence of such time-varying effects one may end up with incorrect models coupled with biased conclusions as a result of misleading effect estimates. Appropriate modeling of the shapes of the covariates is very important since 'incorrect' shapes of the time varying effects could result to misleading conclusions just as erroneously assuming the proportional hazard. Previous studies have shown varying tests and models for the time-varying effects. Cox (1972) proposed a transformation of time which formed a basis for testing and assessing the non-PH, a method that heavily relies on the choice of the time transformation. In this paper we proposed the use of penalized splines in order to disclose and model effects of survival data within the context of cox model framework. The model allows for easy testing of time variation in the presence of effects using standard methods such as likelihood ratio test. However, although the penalized splines (PS) provide a flexible fit, they still suffer from the same restrictions that affect other non-linear smooth functions such as Fractional Polynomials.

## References

[1] Aalen. O. O. (1989). A linear regression model for the analysis of life times. Statistics in Medicine 8 (8) , 907–25.

[2] Aalen. O. O. (1980). A model for non-parametric regression analysis of counting processes. Lecture Notes on Statistics 2 .

[3] Aalen. O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. Statistics in Medicine 12 (17), , 1569–88.

[4] Abrahamowicz, M., , T. MacKenzie,, & J. M. Esdaile. (1996). Time-dependent hazard ratio: Modeling and hypothesis testing with application in lupus nephritis. Journal of the American Statistical Association 91, , 1432–1439.

[5] Berger, U., , J. Schäfer, , & K. Ulm. (2003). Dynamic Cox modelling based on fractional polynomials: Time–variations in gastric cancer prognosis. . Statistics in Medicine 22(7), , 1163–1180.

[6] Boor, d. C. (1978). A Practical Guide to Splines. Springer-Verlag: New-York.

[7] Brown, D., , G. Kauermann, , & I. Ford. (2007). A partial likelihood approach to smooth estimation of dynamic covariate effects using penalised splines. . Biometrical Journal 49, , 1–12.

[8] Buchholz, A. (2010). Assessment of time–varying long–term effects of therapies and prognostic factors.

[9] Cai, Z. , & Y. Sun. (2003). Local linear estimation for time-dependent coefficients in cox's regression model. . Scandinavian Journal of Statistics 30(1), , 93–111.

[10]Coradini, D., et.al. (2000). Time-dependent relevance of steroid receptors in breast cancer. Journal of Clinical Oncology 18, , 2702–2709.

[11]Cox, D. R. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B: Methodological 34, , 187–220.

[12]Eilers, P. H. C. , & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. Stat. Science 11 (2), , 89-121.

[13]Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics 19: 1. .

[14]Gao C, , Yang M,, Wu Y, , & et al. (2006). Hybrid coronary revascularization by endoscopic robotic coronary artery bypass grafting on beating heart and stent placement. Ann Thorac Surg, 87 , 737-41.

[15]Gore, S.M, Pocock, S.J., & Kerr, G.R. (1984). Regression models and non-proportional hazards in the analysis of breast cancer survival. . Applied Statistics, 33(2), , 176–195.

[16]Gray. (1994). Spline based tests in survival analysis. Biometrics , 640–652.

[17]Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis. . Journal of the American Statistical Association 87, , 942–951.

[18]Green, D. J. , & Silverman, B. W. (1994). onparametric Regression and generalized linear models. N Chapman & Hall.

[19]Hastie, T, & Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazard model. Biometrics 46 , 1005-1016.

[20]Hastie, T. , & Tibshirani, R. (1993). Varying coefficient models. Journal of the Royal Statistical Society, Series B 55, , 757-796.

[21]Heinzl, H. , & A. Kaider. (1997). Gaining more flexibility in cox proportional hazards regression models with cubic spline functions. Computer Methods and Programs in Biomedicine 54(3), , 201–208.

[22]Hess, K. R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. Statistics in Medicine 13, , 1045–1062.

[23]Hilsenbeck, S. G., et.al. (1998). Time-dependence of hazard ratios for prognostic factors in primary breast cancer. Breast Cancer Research and Treatment 52, , 227–237.

[24]Lin, D. Y. , & Wei, L. J. (1991). Goodness-of-fit tests for the general Cox regression model. Statistica Sinica 1, , 1-17.

[25]Long J, , & Ryoo J. (2010). Using fractional polynomials to model non- linear trends in longitudinal data. Br J Math Stat Psychol 2010; 63: , 177-203.

[26]Martinussen, T., , T. H. Scheike, , & I. M. Skovgaar. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. Scandinavian Journal of Statistics 29(1), , 57–74.

[27]Ng'Andu, N. H. (1997). An empirical comparison of statistical tests for assessing the proportional hazards assumption of cox's model. Statistics in Medicine 16, , 611–626.

[28]P. M. Grambsch, & T. M. Therneau. (1994). Proportional hazards tests and diagnostics based on weighted residuals. Biometrika, 81 , 515-526, 1994.

[29]Royston, P., & Altman, D. (1994). Regression using Fractional Polynomial of continuous covariates: Parsimonious parametric modeling. Applied Statistics 43, , pp. 429-467.

[30]Schemper, M., S. Wakounig, & G. Heinze . (2009). The estimation of average hazard ratios by weighted cox regression. . Statistics in Medicine 28, , 2473–2489.

[31]Schonfeld, D. . (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. Biometrika 67, , 145-53.

[32]Schumaker, L. (1981). Spline Functions. John Wiley, New York.

[33]Stone, C. (1997). Additive regression and other non-parametric models. . Ann. Statist. , 13, , 689-705.

[34]Therneau, T. M. , & P. M. Grambsch. (2000). Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag Inc.

[35]Verweij, P. J., & Houwelingen, H. C. (1995). Time-dependent effects of fixed covariates in cox regression. Biometrics 51, , 1550–1556.

[36]Wahba, G. (1990). Spline Models for Observational Data. SIAM, Philadelphia .

[37]Yuedong. W. (2011). Smoothing Splines Methods and Applications. New York: Monographs on Statistics and Applied Probability 121.

**Author Profile**

**Albert OtienoOrwa**received a BSc Applied Statistics with Computing from Moi University (Kenya) in 2011.Currently he is studying towards Master of Science in Mathematics (Statistics option)at Pan African University Institute of Basic Sciences, Technology and Innovation (PAUISTI) at Jomo Kenyatta University of Agriculture and Technology (JKUAT) in Kenya.

**Dr George OtienoOrwa**is the current chairman of the Department of Statistics and Actuarial Science at Jomo Kenyatta University of Agriculture and Technology. He has over 11 publications in statistics.

**Professor RomanusOtienoOdhiambo**is a professor in statistics. He received Bachelor of Education, MSc (Statistics), PhD (Statistics) from Kenyatta University. He has published over 25 papers and he is the current Deputy Vice Chancellor (Academic Affairs) at Jomo Kenyatta University of Agriculture and Technology.