SURVIVAL ANALYSIS WITH MULTIVARIATE ADAPTIVE REGRESSION SPLINES USING COX-SNELL RESIDUAL

K. A. Dauda [1], **W. B. Yahya** [2] , **A. W. Banjoko** [3]

[1] **Department of Statistics and Mathematical Sciences, Kwara State University, Malete, P.M.B 1530, Ilorin, Nigeria**
[2,3] **Department of Statistics, Faculty of Physical Sciences, University of Ilorin, P.M.B 1515, Ilorin, Nigeria**

Corresponding author: W. B. Yahya, dr.yah2009@gmail.com

ABSTRACT: Multivariate Adaptive Regression Splines (MARS) are a generalization of stepwise linear regression method that is often employed to improve the efficiency of regression models. It is a useful tool to identify linear/nonlinear and interactions effects between a set of metrical and categorical covariates in regression models. In this study, the use of a modified Cox-Snell Residuals to Survival Analysis with MARS was proposed. The proposed method was compared with Martingale Residual in the Survival MARS setting. These two residual types were used as responses in the Cox proportional hazard modeling in the MARS implementations. Results from simulation studies revealed that the proposed method fitted the data better than the Martingale residual ($\sigma_{cox-snell} = 0.5747 < \sigma_{Martingale} = 0.8356$). However, further results from Monte-Carlo experiment showed that the two residual types performed better than the classical Cox Proportional Hazard (CPH) method. These methods were applied on real life dataset on Pneumocystis Carinii Pneumonia and all the results obtained actually validated those got from the simulation studies.
KEYWORDS: MARS, Martingale Residual, Modified Cox-Snell Residual, PCP and CART.

1. INTRODUCTION

Multivariate adaptive regression splines (MARS) are a useful tool to identify linear and nonlinear effects and interactions between two or more covariates in classical regression models ([B+84]). This approach has been embraced to the area of survival by few researchers. The use of standard analysis of covariate effects such as Cox proportional hazard function can be misleading and difficult to interpret if we have both continuous and binary predictors in the model at the same time. Due to the limitations of these standard methods of analyzing survival models, the MARS technique was embraced as an alternative means of analyzing survival data ([Kri07]).
In contrast to the global parametric modeling technique of Linear Regression Model (LRM), the MARS technique developed by Friedman (1991) is a nonparametric modeling approach of the form $y = f(x) + \epsilon$, where y is a response variable, $x = (x_1, x_2, \dots, x_p)^T$, is a vector of p predictors and ϵ is an additive stochastic component which is assumed to have zero mean and finite variance ([W+12]).
In recent years, MARS has been successfully applied to many areas of science and technology. These include predicting object-oriented software maintainability by Zhou et al. ([ZL07]), waste water treatment by Tsai ([TC05]) and species distributions from presence-only data by Elith et al. ([EL07]). It also has applications in speech modeling, credit scoring, breast cancer pattern determination, and examining the impact of information technology investment on productivity ([W+ 12]).
Among the recent applications of MARS to survival analysis was that of Kriner ([Kri07]) in which MARS was applied to Survival Analysis through the use of both martingale and deviance residuals. In the present work, interest is set on using the Cox-Snell residual to model Survival data with MARS.

2. SURVIVAL ANALYSIS

Survival analysis is basically concerned about waiting time until a particular event of interest occurs. This may be a terminal incident such has death, relapse and so on, but more generally, the event must be a well-defined circumstance ([YU09; YD15]). Survival analysis is sometimes referred to as part of statistical theory that focuses on examining the lifetimes. The waiting time can be discrete or continuous random variables. In this work, emphasis was more on the continuous time random variable.

2.1. Continuous Lifetimes

Let T be the random variable representing the lifetime under study. The distribution function F and the survivor function S of T are defined by the following probabilities:

$$F(t) = P(T \leq t); \; S(t) = P(T > t) \quad (1)$$

Hence, $F(t) + S(t) = 1$ for all t. It should be noted that $F(t)$ is an increasing and $S(t)$ a decreasing, function of t .Normally, $F(t)$ will rise from 0 to 1, and $S(t)$ will fall from 1 to 0, over the range of t. The density function $f(t)$ is defined as

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad (2)$$

Correspondingly, $F(t) = \int_0^t f(s)ds$ and $S(t) = \int_t^\infty f(s)\,ds$ ([Cro12]).

2.2. The Hazard Function

The hazard function is defined as

$$h(t) = \lim_{\delta \downarrow 0} \delta^{-1} P\left(T \leq t + \frac{\delta}{T} > t\right)$$

The right-hand side can be expressed as

$$\lim_{\delta \downarrow 0} \delta^{-1} P\left(T \leq t + \frac{\delta}{T} > t\right) = \frac{\lim_{\delta \downarrow 0} \delta^{-1}\{S(t) - S(t+\delta)\}}{S(t)} = -\left(\frac{dS(t)}{dt}\right)\Big/_{S(t)} = -\frac{d\log S(t)}{dt}$$

In different contexts $h(t)$ is variously known as the instantaneous failure rate, age-specific failure rate, age-specific death rate, intensity function, and force of mortality or decrement ([YU09; Cro12]).

2.3. Some Continuous Survival Distributions

A number of distributions are known to be suitable to model survival times depending on the situations and preference by the users. The most common among these are the Exponential, Gompertz, Log-logistic, Gamma, Normal and Log-Normal distributions. In this study however, interests were focused on the Exponential, Weibull and Gompertz distributions. The reason for this choice lies in the fact that these three distributions, unlike some of their other parametric counterparts, shared the assumption of proportional hazards of the Cox regression model ([Lee92]). Further details on these distributions are provided in what follows within the context of usage in this work.

2.3.1. The exponential distribution

The exponential distribution is the most frequently used distribution in survival analysis. This is mainly due to its flexibility. Lifetimes (T) that are exponentially distributed have density function defined by

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & for\ t > 0, \lambda > 0 \\ 0, & otherwise \end{cases} \quad (3)$$

with $E(T) = \frac{1}{\lambda}$ and $Var(T) = \frac{1}{\lambda^2}$. If $\lambda = 1$, we say that the variables are unit exponentially distributed.

The reliability function $R(t)$ which is the probability that a device or system will perform its intended function for a given interval of time under specified operating conditions is defined by

$$R(t) = e^{-\lambda t}, \quad for\ t > 0 \quad (4)$$

One of the main things to notice about the exponential distribution is that the hazard rate function is constant

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

This again leads to the memory less property of the exponential distribution. The conditional reliability function is defined as the probability that an item will survive for an additional time x, given that it has survived up to t:

$$R\left({}^{x}/_{t}\right) = P\left({}^{T > t + x}/_{T > t}\right) = \frac{P(T > t + x)}{P(T > t)} \rightarrow R\left({}^{x}/_{t}\right) = \frac{\lambda e^{-\lambda(t+x)}}{\lambda e^{-\lambda t}} = e^{-\lambda x} \quad (5)$$

Thus, the relationship $P(T > x) = R(x)$ is trivial from the above.

The consequence of result (5) is that an old item is as good as a new item. The memory less property is a reason for why the exponential distribution is so easy to work with, but this is also its weakness since it is not realistic that an item will have the same hazard function throughout its lifetime ([RYO15]).

2.3.2. The Weibull distribution

The Weibull distribution, named after the Swedish professor Waloddi Weibull, is also one of the most used distributions in survival analysis. It has the density function given by

$$f(t) = \begin{cases} \alpha\lambda^{\alpha}t^{\alpha-1}e^{-(\lambda t)^{\alpha}}, for\ t > 0, \lambda > 0 \\ 0, \quad if\ otherwise \end{cases} \quad (6)$$

with $E(T) = \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha}+1\right)$ and

$$Var(T) = \frac{1}{\lambda^2}\left\{\Gamma\left(\frac{2}{\alpha}+1\right) - \left(\Gamma\left(\frac{1}{\alpha}+1\right)\right)^2\right\}.$$

Here, α is the shape parameter, λ is the scale parameter and $\Gamma(.)$ is the gamma function. The reliability function of the Weibull distribution is defined by

$$R(t) = e^{-(\lambda t)^{\alpha}}, for\ t > 0 \quad (7)$$

If $\alpha = 1$, the Weibull distribution in (6) reduces to the exponential distribution in (3). Thus, the exponential distribution is therefore a special case of the Weibull distribution.

2.3.3. The Gompertz distribution

Gompertz distribution is an illustration of extreme values distribution. Let T be a lifetime that is independently and identically distributed. If the probability density function of T goes exponentially towards zero when $t \to \infty$, the limiting distribution of a normalized version of $U_n = T_{(1)} = \min\{T_1, T_2, \dots, T_n\}$ is known to be

$$F_{T_{(1)}}(t) = 1 - e^{-e^{\left(\frac{t-v}{\alpha}\right)}}, \text{ for } -\infty < t < \infty \quad (8)$$

where α > 0 (the mode) and v (the scale parameter) are constants. If the lifetimes are standardized as $Y = \frac{T - v}{\alpha}$, this produces the cumulative density function of the standardized Gompertz distribution of the smallest extreme given by

$$F_{Y_{(1)}}(y) = 1 - e^{-e^{y}}, \text{ for} -\infty < y < \infty \quad (9)$$

which consequently yielded the following density function

$$f_{Y_{(1)}}(y) = e^{y}e^{-e^{y}}, \text{ for } -\infty < y < \infty \quad (10)$$

with $[Y_1] = -\gamma$; $Var[Y_1] = \frac{\pi^2}{6}$ and γ = 0.5772... is the Euler's constant. The reliability function of the Gompertz distribution is defined by

$$R(y) = e^{-e^{y}}, -\infty < y < \infty \quad (11)$$

2.4. The Cox Proportional Hazards Model

The Cox proportional hazard (PH) regression model ([Cox72]) assumes that the covariates multiplicatively shift the baseline hazard function and assert that the hazard rate for the jth subject in the data is

$$h\left({t}/{X_j}\right) = h_0(t)exp(\beta_x X_j) \quad (12)$$

where the regression coefficients β_x are to be estimated from the data ([CGG04]). The baseline hazard represents the hazard of an 'average' subject of the data, thus the hazard of a particular subject is just a multiple of the baseline hazard. In the Cox model the most important assumption is the proportionality of the hazard rate. That is, for any given input factor X_j with two binary levels (0,1), the hazard rate is given by;

$$\frac{h\left({t}/{X_j}\right)}{h\left({t}/{X_j = 0}\right)} = \frac{h_0(t)exp(\beta_x X_j)}{h_0(t)} = exp(\beta_x X_j) \quad (13)$$

independent of time t. This means that the hazard rate of a particular subject remains the same over the whole observation time. That is, the influence of all the covariates in the model is independent of the survival/observational time t ([Kri07; YU09]).

3. THE MODELS' DEVELOPEMENT

The MARS techniques had been so far actualized and utilized for continuous and binary responses. The aim of this work is to extend this method to survival time information in which the Cox-Snell residuals of a Cox PH model shall be used as response variable in a typical MARS approach.

In spite of the semi-parametric nature of the Cox PH model which makes it different from the standard parametric model with less assumption on the survival time, there exists a number of diagnostic test that can be employed to check the model's adequacy, misspecification, outliers and influential points just to mention a few. These diagnostics tests include the test based on re-estimation (the *Link test*), the tests based on residuals (*Schoenfeld Residuals, Cox-Snell Residuals, Martingale Residuals* and *Deviance Residuals*) and the graphical diagnostic methods (see [Hes95] and [Gar97] for more detail).

Of the above residual types of the Cox PH model, the Martingale and Deviance residuals have been successfully employed as response variables in the common MARS implementations ([Kri07]). In this work therefore, a model that used the Cox-Snell Residual for survival analysis with MARS is proposed. Results from this new model were compared with those that employed the Martingale residuals with MARS technique.

3.1. The Cox- Snell (CS) Residual

Basically, the Cox-Snell (CS) residual like its other counterparts is often employed as a measure of goodness-of-fit of the Cox PH model. Under the generalized residual settings, the Cox PH model using the survivor function

$$S(t; X_j) = [S_0(t)]^{exp(\beta_x X_j)}$$

is fitted to the data. In terms of hazard, the model fitted is of the form

$$h(t, X_j) = h_0(t) exp(\beta_x X_j)$$

Therefore, for each person with covariates X_j, the estimated survivor function is

$$\hat{S}(t)(t, X_j) = \left[\left[\hat{S}\right]_0(t)\right]^{exp(\beta_x X_j)} \quad (14)$$

This gives a predicted survival probability at each time t in the dataset.

Then, the CS residual is calculated by

$$\hat{\Lambda}_j = -\log\left[\hat{S}(T_j; X_j)\right] \quad (15)$$

Additionally, CS residual can be used to examine the overall fit of the Cox PH model by computing

$$h_j(t) = \hat{H}_0(t) exp\left(\sum_{k=1}^{p} b_k x_{jk}\right), j = 1, \dots, n$$

where $b_k = (b_1, \dots, b_p)^t$ is the Maximum Likelihood Estimator (MLE) of β and $\hat{H}_0(t)$ is the MLE of the baseline cumulative hazard function ([KM03]).

Within the framework of MARS technique, the whole space of the input variable is divided into k sub-regions or groups. MARS defines a different mathematical equation for each group. This equation relates each sub-region of input variable to the output variable.

The CS residual is then computed using the following quantity;

$$\widehat{CS}^1_{ik} = I_{\{\delta_{ik=1}\}} - \hat{\Lambda}^1_0(T_{ik}) \quad (16)$$

where $\hat{\Lambda}^1_0(.)$ is the estimated cause-specific cumulative hazard function which is computed by

$$\hat{\Lambda}^1_0(T_{ik}) = \int_0^{T_{ik}} \frac{dN_{1k}(S)}{Y_k(s)} \quad (17)$$

where $N_{1k}(t)$ and $Y_k(t)$ are the number of the event observed and the number of subjects at risks at time t in group k respectively.

MARS uses the following two-sided truncated power functions as splines basis functions:

$$[-(x-t)]^q_+ = \begin{cases} (t-x)^q, if\ x < t \\ 0, \quad otherwise \end{cases}$$
$$[+(x-t)]^q_+ = \begin{cases} (x-t)^q, if\ x \geq t \\ 0, \quad otherwise \end{cases}$$

where q is the power and t is the number of knots. This gives rise to the final model

$$\hat{Y} = \hat{f}(x) = a_0 + \sum_{m=1}^{m} a_m B_m(x) \quad (18)$$

where $\hat{Y}$ is the output variable, x is the input variable, a_0 is the constant term, m is the number of the splines functions, and a_m and B_m are the m^{th} splines function and its coefficient respectively.

3.1.1. The MARS Algorithm

The execution of the MARS technique follows the following algorithm as adapted from Hastie et al. ([HTF01]):

i. **Stage-wise forward selection**: Keep the coefficient the same for variable that existed in the current model, select a new basis function pair that produces the largest decrease in the training error. Repeat the whole process until some termination condition is met, for instance:
   a. reached the maximum number (pre-determined) of predictors;

b. adding any new predictor changes the R-Square by less than 0.001;
c. Reached a very high R-Square value.

ii. **Backward Algorithm:** For preventing over fitting, redundant basis functions are deleted from the model. MARS adopts Generalized Cross-Validation (GCV) to delete the redundant basis functions. The expression of GCV is given by:

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N}[Y_i - \hat{f}(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2}$$

where N is the number of data points and $C(B)$ is the complexity penalty that increases with the number of basis function (BF) in the model and is defined as: $C(B) = (B+1) + dB$, where d is a penalty for each BF included in the model and B is the number of the basis functions in the model.

3.2. The Martingale Residuals

Martingale residuals are used to examine the functional form of a given covariate in order to make the best use of it to explain its influence on the Cox PH model. If the test tells us that we could not introduce a covariate into the model linearly, we might need to transform the covariate X into $log(X), exp(X), XlogX$ or even discretized version of the covariate. Suppose we have a vector $X_j(t)$ of possible time-dependent covariates. Let $N_j(t)$ have a value equals to 1 at time t if this individual has experienced the event of interest and 0 if the individual has yet to experience the event of interest. Let $Y_j(t)$ be the indicator that individual j is under study at a time just prior to time t. Finally, let b be a vector of regression coefficients and $\widehat{H}_0(t)$ be the Breslow estimator of the cumulative baseline hazard rate ([Bre72]). Then, the martingale residual is defined ([M03]) as

$$\widehat{M}_j = N_j(\infty) - \int_0^\infty Y_j(t)exp(b'X_j(T))d\,\widehat{H}_0(t), \quad j = 1, \dots, n. \tag{19}$$

When the data is right-censored and all the covariates are time-independent, the martingale residual reduces to

$$\widehat{M}_j = \delta_j - \widehat{H}(T_j)exp\left(\sum_{k=1}^{p} X_{jk}\, b_k\right) \tag{20}$$

$$\widehat{M}_j = \delta_j - r_j, j = 1, \dots, n$$

where δ_j is indicator of censoring for j individual and r_j is the corresponding Cox-Snell residual.

3.3. The Deviance Residuals

Martingale residuals are defined on the interval (−∞, 1] and thus is highly skewed. To determine whether this Skewness does in any way affect the outcome, the new method was also applied to a transformation of the martingale residuals, namely deviance residuals, which are symmetrically defined on the interval (−∞, ∞). If all covariates are time-fixed, the deviance residual for subject j is defined as

$$\widehat{D}_j = sign\left(\widehat{M}_j\right)\sqrt{\left(-\widehat{M}_j - N_j \log\left(\frac{N_j - \widehat{M}_j}{N_j}\right)\right)} \tag{21}$$

4. SIMULATION STUDIES

Simulation studies are an important statistical tool often employed to examine the performance, power and other properties of a new statistical method or approach. Thus, various schemes employed for the Monte-Carlo experiment in this study were presented in this section.

4.1. Simulating Survival Times

Survival times were simulated from the three selected distributions; i.e. Exponential, Weibull and Gompertz distributions - as earlier remarked in Section 2.3. Table 1 shows the methods employed for generating the survival times across these three distributions. Various results from these implementations were provided in the Appendices A, B and C.

Some general considerations employed in the Monte-Carlo experiment are based on the following: The survival function of the Cox PH model is given

$$S(t/x) = exp\left[-H_0(t)exp\left(f(x)\right)\right]$$

and the cumulative survival function is defined as

$$F(t/x) = 1 - S(t/x)$$

where $H_0(t) = \int_0^t h_0(u)du$, is the cumulative baseline hazard function ([YU09]).

Table 1: Statistics for generating survival times across the specified distributions

	Distribution		
	Exponential	**Weibull**	**Gompertz**
Parameter	Scale parameter $\lambda > 0$	Scale parameter $\lambda > 0$ Shape parameter $V > 0$	Scale parameter $\lambda > 0$ Shape parameter $\alpha \in (-\infty, \infty)$
Survival time	$T = -\frac{\log(u)}{\lambda \exp(f(x))}$	$T = \left[-\frac{\log(u)}{\lambda \exp(f(x))}\right]^{\frac{1}{v}}$	$T = \frac{1}{\alpha}\log\left[1 - \frac{\alpha \log(u)}{\lambda \exp(f(x))}\right]$
Hazard function	$h(t) = \lambda \exp(f(x))$	$h(t) = \lambda \exp(f(x))\, v t^{v-1}$	$h(t) = \lambda \exp(f(x)) \exp(\alpha t)$

Five covariates $x_1, \dots, x_5$ were simulated over n = 1000 samples from different distributions as follows:

$x_1 \sim U[30,75]$, $x_2 \sim bin(n, 0.5)$, $x_3 \sim N(0,1)$, $x_4 \sim U[0,10]$ and $x_5 \sim bin(n, 0.7)$ with $u \sim U[0,1]$.

In the simulation of survival times (t) under the three distributions (Exponential, Weibull and Gompertz) employed in this work, only the linear, nonlinear and interaction effects of covariates x_1 and x_2 on the response variables (t) were considered. The other covariates x_3, x_4 and x_5 were introduced into the model as nuisance covariates that are expected to be identified by the survival MARS model itself. This concept was first introduced by Friedman ([Fri91]) and later by Kriner ([Kri07]).

Therefore, for the Exponential distribution, the survival time t was simulated according to the scheme adopted by Kriner ([Kri07]) using the relation:

$$t = \frac{-\log(u)}{0.5 \exp\left[0.04\left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right]}$$

For the Weibull distribution, the survival time t was simulated from the relation

$$t = \left[\frac{-\log(u)}{\lambda \exp\left[0.04\left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x1x2\right]}\right]^{\frac{1}{v}}$$

with the shape parameter $v = 2$ and the scale parameter $\lambda = 5$ (See [Kri07]).

For the Gompertz distribution, the density and survival functions of this extreme value distribution with parameters λ and α are given by

$$f_0(t) = \lambda \exp(\alpha t) \exp\left(-\frac{\lambda}{\alpha}\exp(\alpha t)\right) \quad (22)$$

$$S_0(t) = \exp\left(-\frac{\lambda}{\alpha}\exp(\alpha t)\right) \quad (23)$$

with the expressions for the mean and variance given by;

$$E(T) = \mu_0 = -\frac{1}{\alpha}\left(\log\left(\frac{\lambda}{\alpha}\right) + \gamma\right) \quad (24)$$

$$V(T) = \sigma_0^2 = \frac{\pi^2}{6\alpha^2} \quad (25)$$

where $\gamma \approx$ 0.5772 is the Euler's constant and $\pi = 3.14159$. Solving equation (24) and (25) for λ and α gives $\alpha = \frac{\pi}{\sqrt{6}\sigma_0}$, $\lambda = \alpha \exp(-\gamma - \alpha\mu_0)$.

The values of α, λ and γ determined above were used to compute the approximate parameters of the Gompertz distribution independence on the mean and variance of the considered survival time. The adopted scheme is similar to that of Bender et al. (2005) by setting the mean life expectancy μ_0 at $\mu_0 = 66.86$ years with a standard deviation of $\sigma_0 = 6$ years. By this, the values of λ and α were explicitly determined to be λ = 7×10^{-8} and α = 0.2138. Hence, the survival time t was simulated from the relation

$$t = \frac{1}{0.2138}\log\left[1 - \frac{0.2138\log(u)}{0.0000007\exp\left(0.04\left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right)}\right]$$

5. SIMULATION RESULTS

In the Monte-Carlo experiment, 1000 simulation cycles were computed for four different values of the penalizing parameter d , $d{:}\in\{2,3,4,5\}$, with martingale and Cox-Snell residuals as responses. Also, three different censoring rates of low (25%), moderate (50%) and high (85%) were considered in the analysis.

At each fit of survival MARS model on the simulated data, the amount of non-linear influence of the covariate x_1 detected by the models (expressed in %) as well as that of its interaction with x_2 at different values of the penalizing parameters were determined under the Cox-Snell and Martingale residual responses.

Also, the amount of false basis function (in %) chosen by each model was also determined under the two residual types.

Tables 2, 3 and 4 presented the results of the simulation studies under the three selected distributions employed for simulating the survival times.

The implications of the results in Table 2 regarding the behavior of predictor variables x_1 and x_2 in the proposed (Cox-Snell) and the existing (Martingale) models under the Exponential survival times were clearly shown by the graphs (Figs. 1 to 4) in Appendix A at the selected values of the penalty (tuning) parameter d = 2 and 5 with low (25%) and high (85%) censoring rates.

It was found that both models identified the inherent linear relationship of the nuisance variables x_3, x_4 and x_5 with the response variable (t). However, structure displayed by the various graphs showed that the survival MARS model with Cox-Snell residual (proposed model) was able to capture the linear, nonlinear and interaction effects of the covariates x_1 and x_2 better than the survival MARS model with Martingale residuals (existing model) (see Figs. 1 to 4).

Table 2: Results of Survival MARS with Exponential distributed survival time. The amount of non-linear effects of covariates x_1 (in %) and its interactions with x_2 with Martingale and Cox-Snell residuals as responses were reported

d	Martingale Residuals				Cox-Snell Residuals			
	Censoring rate	**%knots(x_1)**	**% x_1x_2**	**%false basis**	**Censoring rate**	**%knots(x_1)**	**% x_1x_2**	**%false basis**
2	25%	70%	100%	50%	25%	**80%**	100%	50%
	50%	100%	100%	0%	50%	100%	100%	30%
	85%	100%	80%	70%	85%	90%	30%	80%
3	25%	70%	100%	40%	25%	80%	100%	40%
	50%	100%	100%	0.0%	50%	100%	100%	0.0%
	85%	100%	50%	60%	85%	100%	100%	20%
4	25%	40%	100%	60%	25%	60%	100%	50%
	50%	100%	100%	0.0%	50%	100%	90%	30%
	85%	100%	80%	30%	85%	50%	0.00%	100%
5	25%	20%	100%	80%	25%	50%	100%	50%
	50%	90%	100%	10%	50%	100%	100%	0.0%
	85%	100%	30%	70%	85%	50%	0.0%	100%

Table 3: Results of the simulation studies on Weibull distributed ($v = 2, \lambda = 5$) survival time with Martingale and Cox-Snell residuals

d	Martingale Residuals				Cox-Snell Residuals			
	Censoring rate	**%knots(x_1)**	**% x_1x_2**	**%false basis**	**Censoring rate**	**%knots(x_1)**	**%x_1x_2**	**%false basis**
2	25%	80%	100%	50%	25%	100%	100%	40%
	50%	100%	100%	0%	50%	100%	100%	50%
	85%	100%	50%	70%	85%	100%	50%	60%
3	25%	40%	100%	60%	25%	100%	100%	0%
	50%	100%	100%	10%	50%	100%	90%	30%
	85%	100%	50%	60%	85%	100%	60%	60%
4	25%	10%	100%	90%	25%	100%	100%	50%
	50%	100%	90%	20%	50%	100%	80%	50%
	85%	90%	30%	80%	85%	90%	40%	50%
5	25%	0%	100%	100%	25%	100%	100%	0%
	50%	100%	100%	30%	50%	100%	60%	40%
	85%	100%	40%	60%	85%	100%	50%	50%

Table 4: Results of the simulation studies on Gompertz distributed survival time with Martingale and Cox-Snell residuals

d	Martingale Residuals				Cox-Snell Residuals			
	Censoring rate	**%knots(x_1)**	**%x_1x_2**	**%false basis**	**Censoring rate**	**%knots(x_1)**	**%x_1x_2**	**%false basis**
2	25%	60%	80%	50%	25%	80%	60%	70%
	50%	90%	30%	80%	50%	90%	50%	70%
	85%	70%	10%	100%	85%	90%	30%	80%
3	25%	60%	100%	70%	25%	90%	40%	80%
	50%	60%	60%	70%	50%	80%	10%	90%
	85%	80%	40%	60%	85%	80%	0%	100%
4	25%	40%	80%	80%	25%	70%	30%	100%
	50%	30%	10%	90%	50%	80%	0%	100%
	85%	80%	0%	100%	85%	60%	20%	90%
5	25%	30%	80%	70%	25%	60%	20%	80%
	50%	50%	10%	90%	50%	50%	0%	100%
	85%	50%	0%	10%	85%	40%	0%	100%

Furthermore, the graphical presentation of the results for Weibull distributed survival times was also presented in appendix B (Figs. 5 to 8). These graphical results equally provided information on the implication of the results in Table 3 at penalty parameter values d = 2 and 5 with low (25%) and high (85%) censoring rates. Results from the graphs (Figs 5 to 8) showed that both the Cox-Snell and Martingale models identified the nuisance variables x_3, x_4 and x_5 to be linearly concomitant to the response variable (t), but the interaction effect of covariates x_1 and x_2 is tighter (under the Cox-Snell model) at large value of the penalty parameter (d = 5) than at low values of d (d = 2) especially at high censoring rate.

Finally, the results from the survival MARS under the two residual types with Gompertz distributed survival times followed similar patterns like those earlier presented. The graphical summary of the non-linear effects of the predictor variables x_1 and x_2 as well as their interactions were clearly shown by the graphs Appendix C (see Figs. 9 to 12). The linear association between other covariates x_3, x_4 and x_5 with survival time was also obvious from the various graphs.

6. APPLICATION TO PNEUMOCYSTIS CARINIIP NEUMONIA DATA

To validate the results from the Monte-Carlo study presented in Section 5, the survival MARS models with Cox-Snell and Martingale residuals were both fitted to a published real life dataset on Pneumocystis Cariniip Neumonia (PCP). These data had been previously analyzed by Lee and Wang ([LW03]). The PCP, a life-threatening disease, has been reported to be the most common opportunistic infection in HIV-infected patients ([LW03]). Many North Americans with AIDS have had one or two episodes of PCP during the course of their HIV infection. The PCP has therefore become a factor of considered in mortality and morbidity modeling while its recurrences are often common.

The data consist of time to recurrence (in months) of PCP with four categorical predictors – gender (male, female), race (white, black, others), Homosexuality practice (yes, no) and two metrical (continuous) covariates - CD4 count and Weight. The outcome (status) variable is categorized as 1 if there was a relapse and 0 if no relapse occurred (censored). Results in Table 5 showed that the Survival MARS model with Cox-Snell residual performed better than the Martingale residual model using the standard errors of the estimated parameters as assessment criteria.

To allow broader comparison of the efficiency of the proposed survival MARS model using Cox-Snell residual with the classical Cox PH model, the Cox PH model was fitted to the data and the results obtained were equally reported in the Table 5. From the results in Table 5, it can be observed that the Cox-Snell residual model was more efficient than the Cox PH model on the real life data as well. In the three cases, the standard errors of the estimated parameters by the Cox-Snell survival MARS model were relatively smaller than those provided by both the Cox PH and Martingale residual models.

Table 5: Results of Survival MARS models using Cox-Snell and Martingale residuals on real life (PCP) data. Results of the classical Cox PH model fitted to these data were also reported in the table

Predictors	Coefficient (β)	Standard Error of β
Survival MARS with Martingale residual		
CD4 count	-0.00014	0.00025
Gender (Female = Ref.)	0.15814	0.10750
Race (white=1) (Other = Ref.)	0.01463	0.05310
Race (black=2) (Other = Ref.)	-0.01044	0.05880
Weight	0.00053	0.00053
Homosexuality (No = Ref.)	-0.01291	0.04868
Survival MARS with Cox-Snell residual		
CD4 count	0.00005	0.00005
Gender (Female = Ref.)	-0.01774	0.02511
Race (white=1) (Other = Ref.)	0.02390	0.01226
Race (black=2) (Other = Ref.)	-0.30309	0.01354
Weight	-0.00006	0.00012
Homosexuality (No = Ref.)	0.00422	0.01133
Classical Survival Analysis (Cox PH Model)		
CD4 count	0.99900	0.00215
Gender (Female = Ref.)	2.85×10^{-7}	3760
Race (white=1) (Other = Ref.)	1.21000	0.7480
Race (black=2) (Other = Ref.)	0.98500	0.84200
Weight	1.00000	0.00434
Homosexuality (No = Ref.)	0.73600	0.37700

7. DISCUSSIONS

The main objective of this work was to examine the efficiency of a new technique for modeling survival MARS with Cox-Snell residuals relative to some of the existing methods (the Cox PH and Martingale residual methods) under three different distributions of survival times – Exponential, Weibull and Gompertz distributions.

Monte-Carlo experiment was carried out to examine the behaviors of some metrical covariate in the two modeling structure in which the Cox-Snell and Martingale residuals were used as responses in the survival MARS models under low, medium and high censoring rates at different values of penalty (tuning) parameters.

Numerical results from the fitted survival MARS models with Cox-Snell and Martingale residuals on the simulated data were presented in Tables 2 (for Exponential survival times), 3 (for Weibull distributed survival times) and 3 (for Gompertz distributed survival times). Graphical summaries of the implication of the results in Tables 3, 4 and 5 were presented by Figs 1 to 12 in Appendices A (for Exponential survival times), B (for Weibull distributed survival times) and C (for Gompertz distributed survival times) respectively.

It was observed from the simulation results in Table 2 that for Exponentially distributed survival times, the proposed Cox-Snell residual model detected the nonlinear influence of x_1 and its interactions with x_2 more than the Martingale residual (existing) model at the penalizing parameter d = 2 and at low (25%) and medium (50%) censoring rates. However, the Martingale residual model detected the nonlinear influence of x_1, its interactions with x_2 and the amount of false chosen basis functions (in %) at the high (85%) censoring rate more than the Cox-Snell residual model.

Further results from Table 2 showed that, when the penalized parameter is 3 and the censoring rates were low, medium and high the Cox-Snell residual model detected the non-linear influence of x_1, its interaction with x_2 and the amount of false chosen basis function than the Martingale residual model.

Without loss of generality, it was found that the efficiency of the Cox-Snell residual model increases in term of its ability to detect the non-linearity of x_1 and its interaction with x_2 as the value of the penalized parameter increases, although at high censoring rate.

Similar results were obtained from the survival MARS models with Cox-Snell and Martingale residuals for Weibull and Gompertz distributed survival times.

Finally, the proposed survival MARS model with Cox-Snell residuals, the survival MARS model with

Martingale residual and the classical Cox PH models were applied on a real life dataset to validate the results from Monte-Carlo studies. The results obtained showed that the proposed Cox-Snell residual model was more efficient than both the existing Martingale residual survival MARS and the classical Cox PH models.

CONCLUSION

In this work, the use of (modified) Cox-Snell Residual to Survival analysis with MARS was proposed. The proposed method was compared with Martingale Residual Survival MARS modeling technique and the classical Cox PH model. Both the modified Cox-Snell and Martingale residuals of a Cox PH model were used as responses in a common survival analysis with MARS.

All the results from simulation studies and real life application reveal that the proposed survival MARS model with Cox-Snell residuals fitted the data more appropriately than either the Martingale residual survival MARS and the classical Cox PH model.

It can therefore be concluded that the use of Cox-Snell residual in Survival analysis modeling with MARS technique has immensely improve the efficiency of the existing technique for modeling time-to-event data irrespective of the distribution of the survival times under consideration.

ACKNOWLEDGMENT

The authors thanked the anonymous reviewers and members of editorial board of this journal for excellent review exercise.

REFERENCES

[Bre72] **N. E. Breslow** - *Contribution to the discussion on the paper by DR Cox, Regression and life tables*. Journal of the Royal Statistical Society, B, 34: 202 – 205, 1972.

[BAB05] **R. Bender, T. Augustin, M. Blettner** - *Generating survival times to simulate Cox proportional hazards models.* Statistics in Medicine, 24:1713–1723, 2005.

[B+84] **L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone** - *Classification and Regression* Trees. Wadsworth, Belmont, 1984.

[Cha03] **T. L. Chap** - *Introduction Biostatistics*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

[Cox72] **D. R. Cox** - *Regression models and life tables (with discussion).* Journal of the Royal Statistical Society - B 1972; 34: 187-220, 1972.

[Cro12] **M. J. Crowder** - *Multivariate Survival Analysis and Competing Risks*. Chapman & Hall/CRC Press, Boca Raton, London New York, 2012.

[CGG04] **M. A. Cleves, W. W. Gould, R. G. Gutierrez** - *An Introduction to Survival Analysis Using Stata*. Stata Press, College Station, Texas, 2004.

[EL07] **J. Elith, J. Leathwick** - *Predicting species distribution from museum and herbarium records using multi-response models fitted with multivariate adaptive regression splines*, Diversity and Distributions, Vol. 13, pp.265-275, 2007.

[Fri72] **J. H. Friedman** - *Multivariate Adaptive Regression Splines*. The Annals of Statistics, Vol. 19, No.1: pp. 1 – 67, 1972.

[Fri91] **J. H. Friedman** - *Multivariate Adaptive Regression Splines*. The Annals of Statistics, Vol. 19, No.1: pp. 1 – 67, 1991.

[Gar97] **J. M. Garret** - *Graphical Assessment of the Cox Model Proportional Hazard Assumption. Stata Technical Bulletin 35:9-14*. In: Stata Bulletin Reprint, Vol.6, 38-44. College Station, TX: Stata Press, 1997.

[Hes95] **K. R. Hess** - *Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression*. Statistics in Medicine, 14: 1707-1723, 1995.

[HT90] **T. Hastie, R. Tibshirani** - *Generalized Additive Models Chapman & Hall*, New York, 1990.

[HTF01] **T. Hastie, R. Tibshirani, J. Friedman** - *The Elements of Statistical Learning: Pre-diction, Inference and Data Mining*. Springer Verlag, New York, 2001.

[HTF03] **T. Hastie, R. Tibshirani, J. H. Friedman** - *The Elements of Statistical Learning*. New York: Springer Verlag, 2003.

[Koo13] **C. Kooperberg** - *Polspline: Polynomial spline routines. R package version 1.1.9*. http://CRAN.R-project.org/package=polspline, 2013.

[Kri07] **M. Kriner** - *Survival Analysis with Multivariate adaptive Regression Splines*, https://edoc.ub.uni-muenchen.de/7369/1/Kriner_Monika.pdf, 2007.

[KM03] **J. P. Klein, M. L. Moeschberger** - *Survival Analysis: Techniques for censored and Truncated Data*. 2d ed. New York: Springer, 2003.

[KM97] **J. P. Klein, M. L. Moeschberger** - *Survival Analysis - Techniques for Censored and Truncated Data*. Springer, New York, 1997.

[Lee92] **E. T. Lee** - *Statistical Methods for Survival Data Analysis*. Wiley: New York, 1992.

[LW03] **E. T. Lee, J. W. Wang** - *Statistical Methods for Survival Data Analysis*. Wiley: New York, 3rd Edition, 2003.

[Rip96] **B. D. Ripley** - *Pattern Recognition and Neural Networks*. Cambridge University Press, 403 pages. 0 -521-46086- 7, 1996.

[RYO15] **M. A. Raheem, W. B. Yahya, K. O. Obisesan** - *A Markov Chain Approach on Pattern of Rainfall Distributions in Uyo Metropolis of Nigeria*. Journal of Environmental Statistics, 7(1):1-13, 2015. http://www.jenvstat.org

[Ste11] **A. Stein** - *Residuals and Functional Form in Accelerated Life Regression Models*. Norwegian University of Science and Technology Department of Mathematical Sciences, M.Sc. thesis, 2011.

[TC05] **J. C. C. Tsai, V. C. P. Chen** - *Flexible and robust implementations of multivariate adaptive regression splines within a wastewater treatment stochastic dynamic pro-gram*, in: Quality and Reliability Engineering International, Vol. 21, pp.689-699, 2005.

[W+12] **G. W. Weber, I. Batmaz, G. Koksal, P. Taylan, F. Yerlikaya-Ozkurt** - *CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization*, Inverse Problems in Science and Engineering, 20, 371-400, 2012.

[Yan12] **J. Yan** - *KMsurv Data sets from Klein and Moeschberger Survival Analysis*. R package version 0.1-5. http://CRAN.R-project.org/package=KMsurv, 2012.

[YD15] **W. B. Yahya, K. A. Dauda** - *Explorative Study of Duration of Marriage In Nigeria: Survival Analysis Approach*, African Journal of Educational Studies in Mathematics and Sciences, (Accepted), 2015. http://www.ajol.info/index.php/ajesms

[YU09] **W. B. Yahya, K. Ulm** - *Survival Analysis of Breast and Small-Cell Lung Cancer Patients using Conditional Logistic Regression Model. Int'l. Jour. of Ecological Economics & Statistics*, ***14***, *(S09): 15-34*. *http://www.ceser.res.in/ijees.html*, 2009.

[ZL07] **Y. Zhou, H. Leung** - *Predicting object-oriented software maintainability using multivariate adaptive regression splines*, Journal of Systems and Software, Vol. 80, pp. 1349-1361, 2007.

[***15] **R Development Core Team** - *R: A language and environment for statistical computing. R foundation for Statistical computing*, Vienna, Austria, URL http://www.r-project.org/, 2015.

Appendix A: Graphs of Survival MARS results for Simulated Exponential Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters

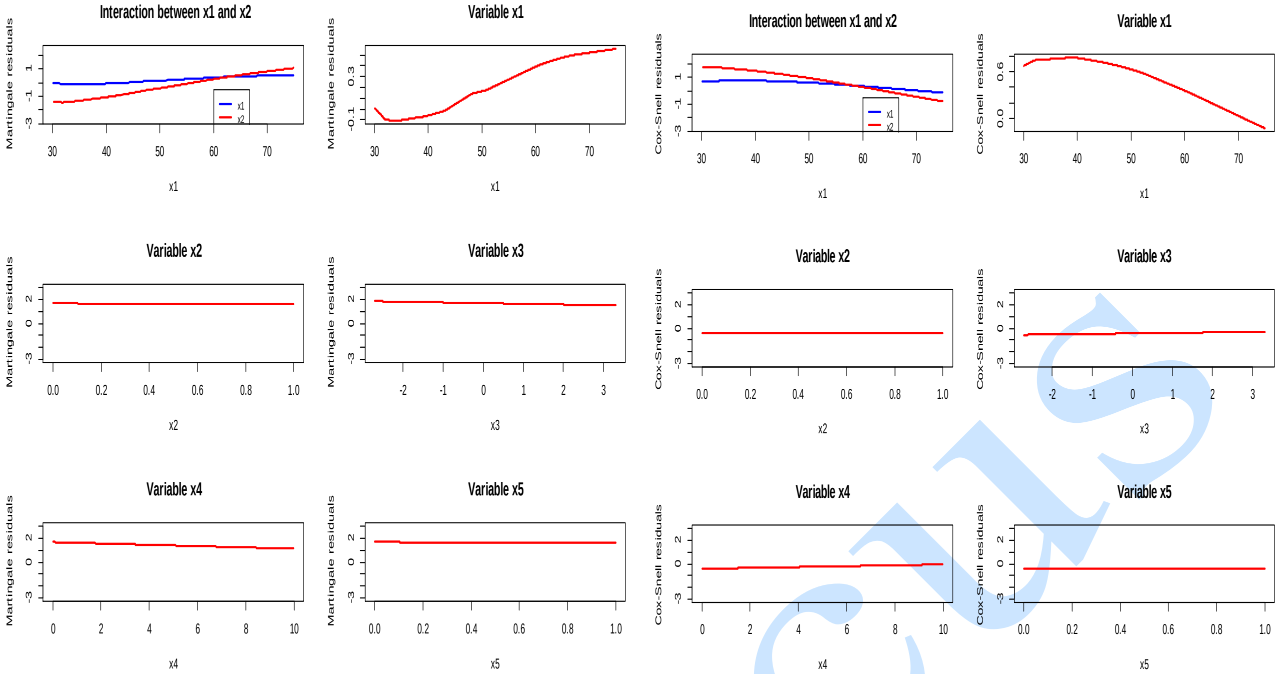

Fig. 1: Graphs of Survival MARS results under Exponential Survival times with penalized parameter $d = 2$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

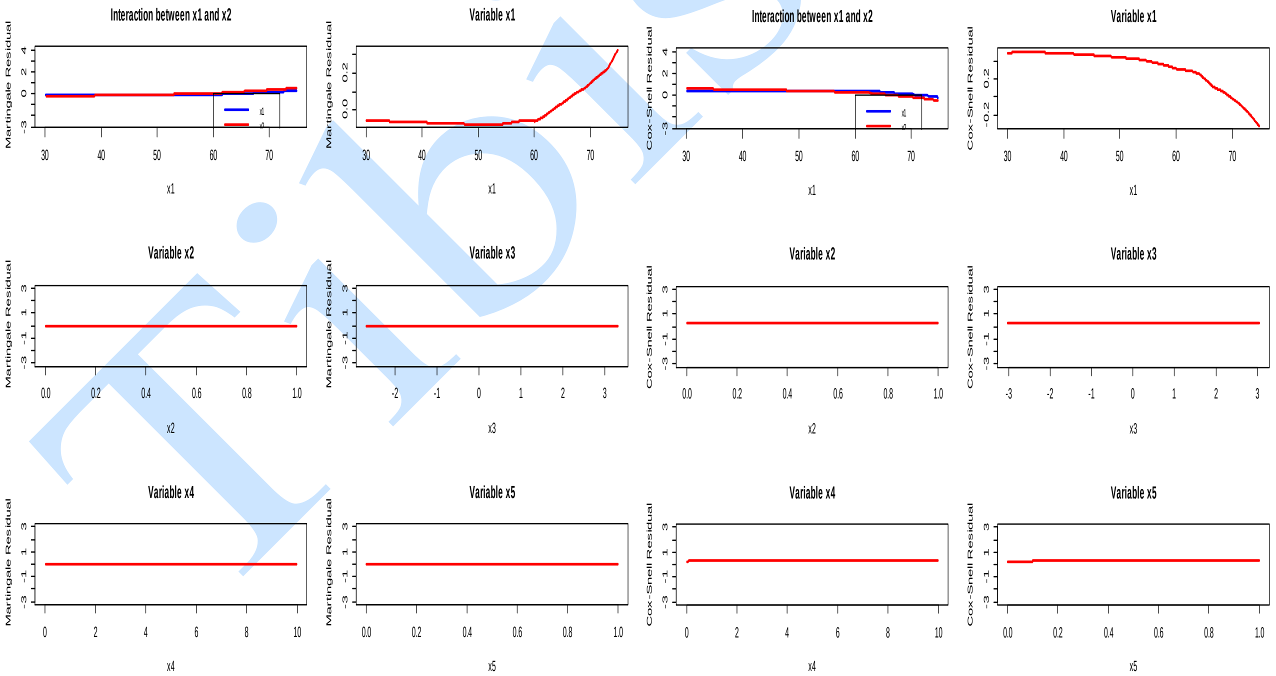

Fig. 2: Graphs of Survival MARS results under Exponential Survival times with penalized parameter $d = 2$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

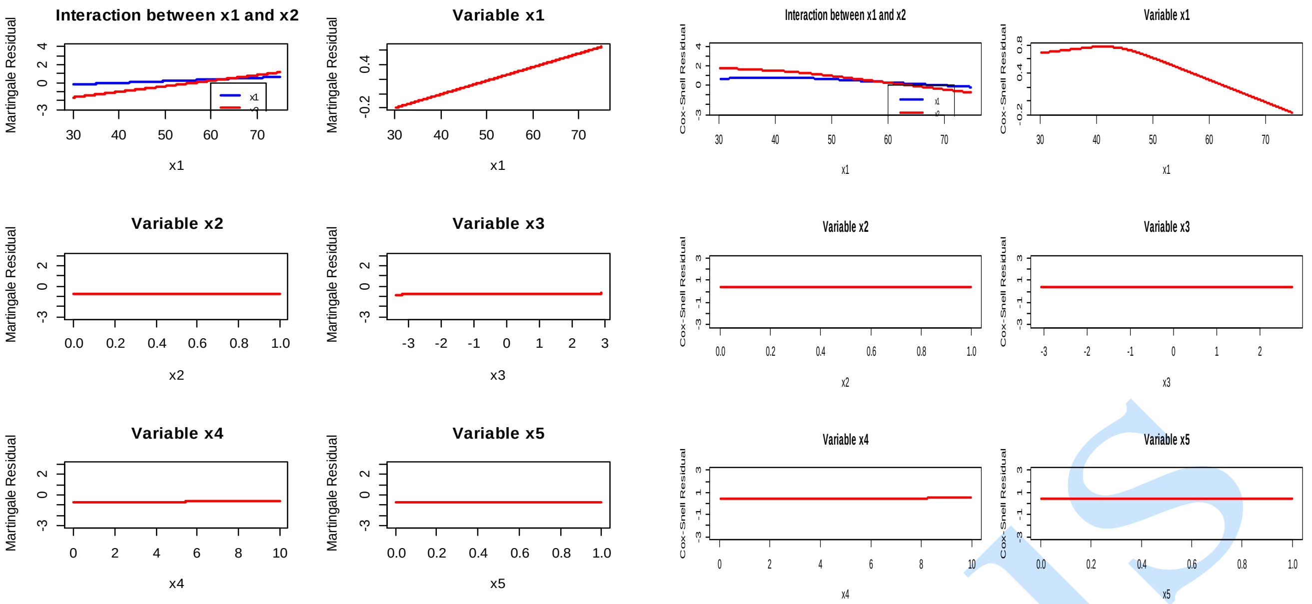

Fig. 3: Graphs of Survival MARS results under Exponential Survival times with penalized parameter $d = 5$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

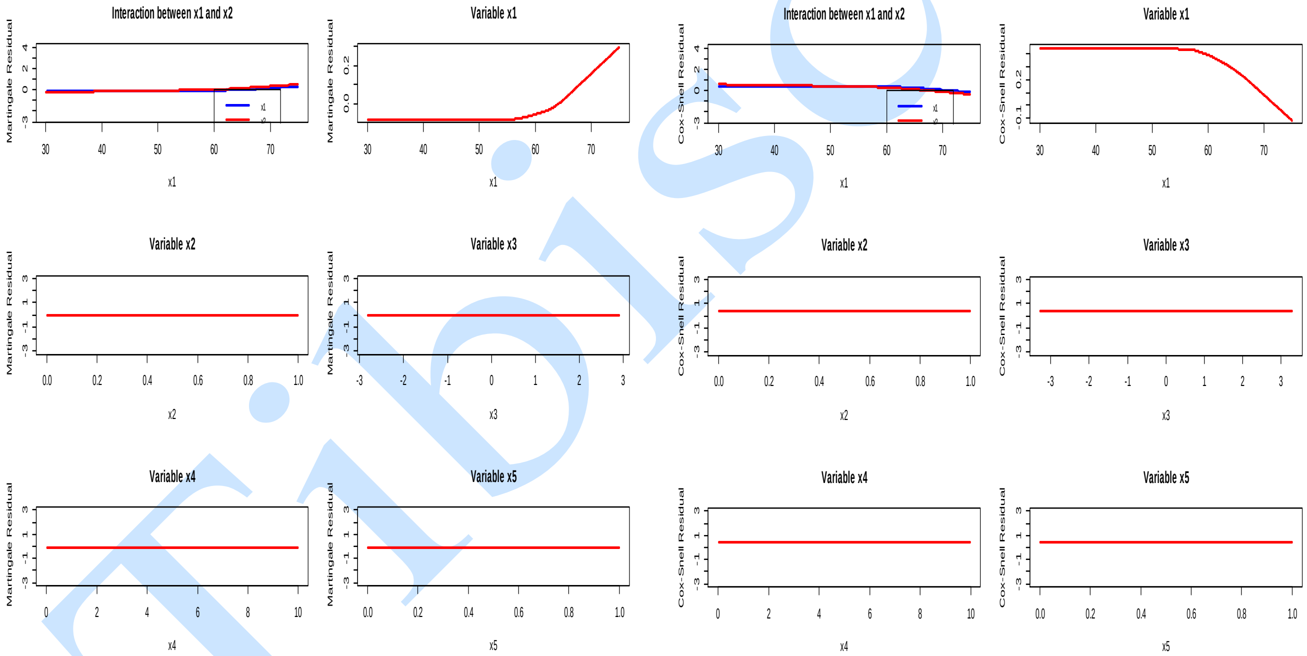

Fig. 4: Graphs of Survival MARS results under Exponential Survival times with penalized parameter $d = 5$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

Appendix B: Graphs of Survival MARS results for Simulated Weibull Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters

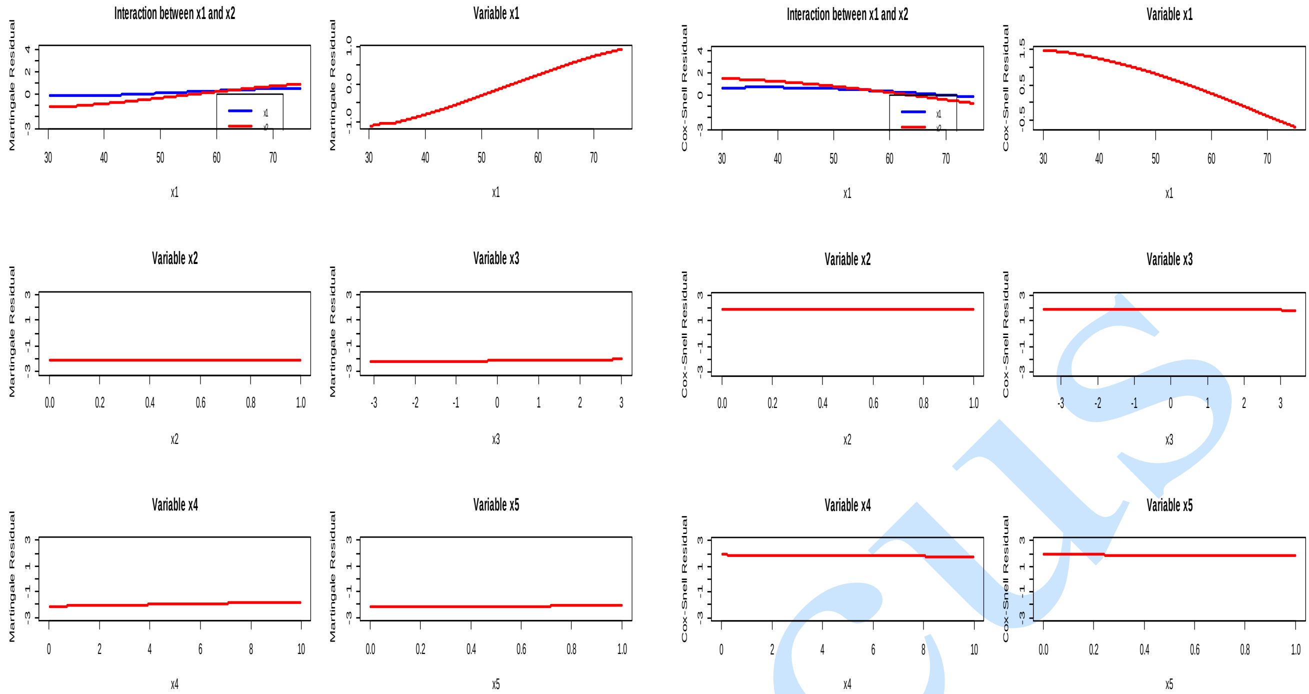

Fig. 5: Graphs of Survival MARS results under Weibull Survival times with penalized parameter $d = 2$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

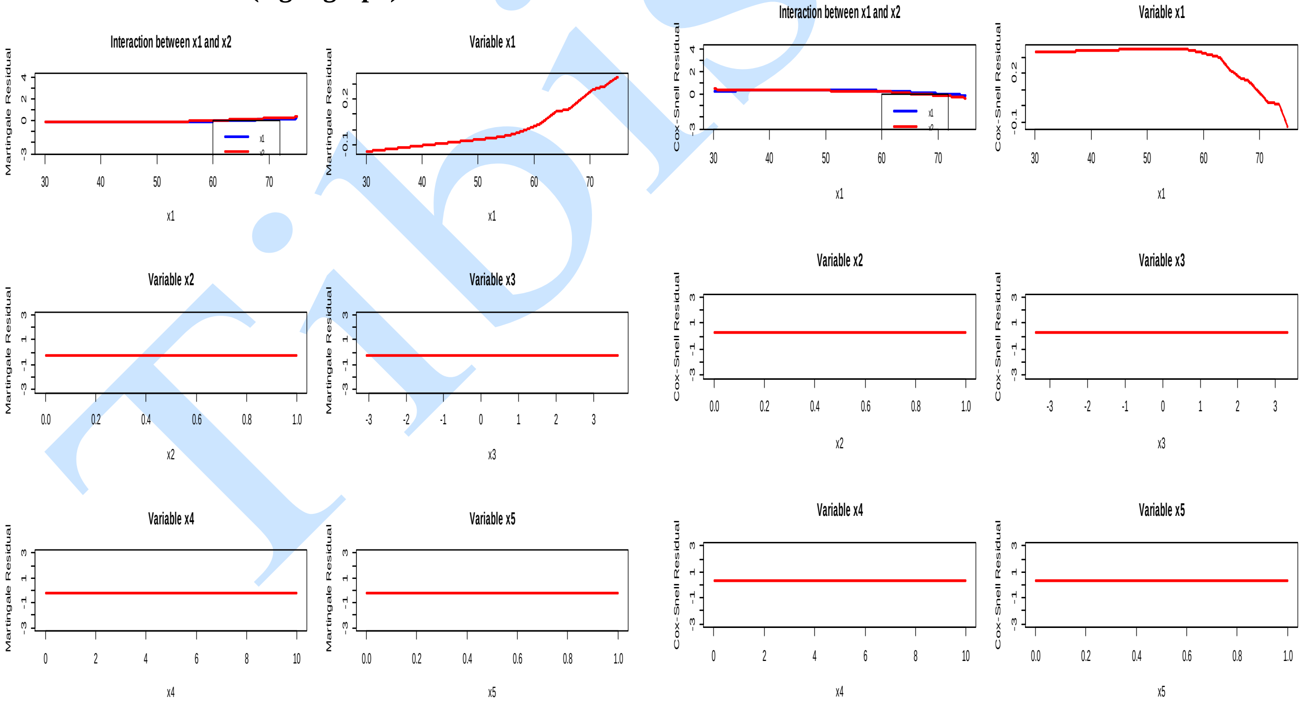

Fig. 6: Graphs of Survival MARS results under Weibull Survival times with penalized parameter $d = 2$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

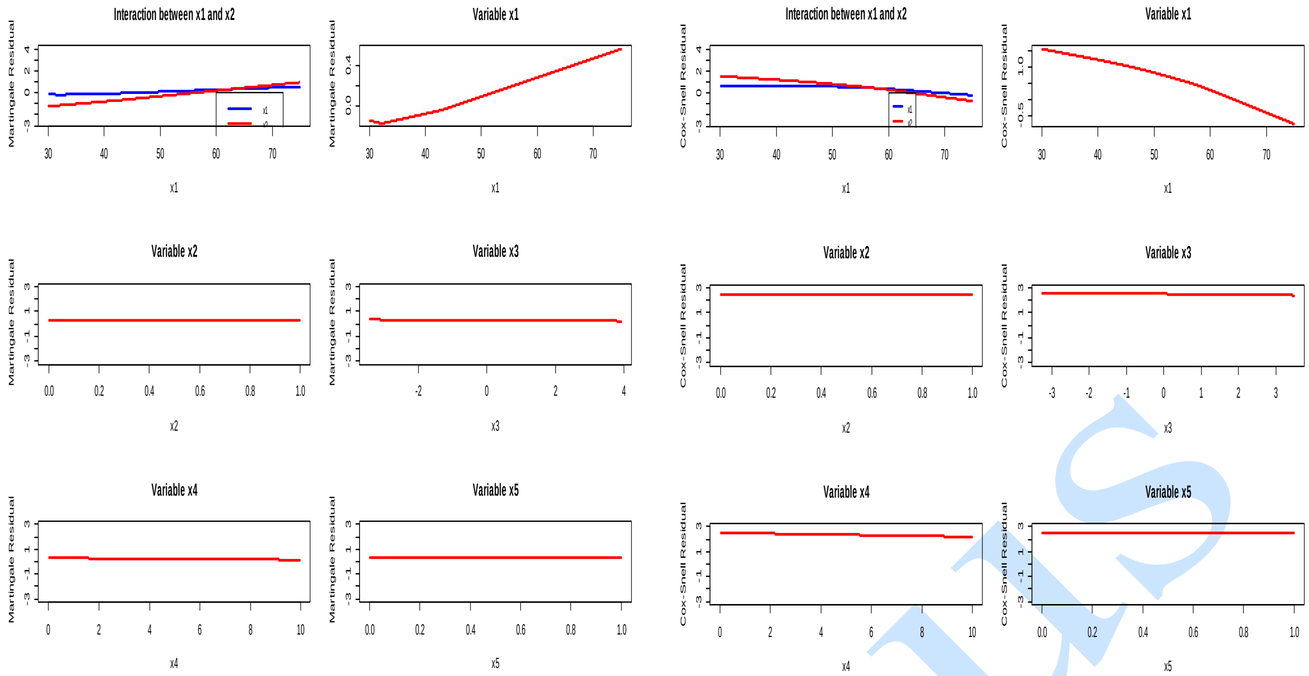

Fig. 7: Graphs of Survival MARS results under Weibull Survival times with penalized parameter $d = 5$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

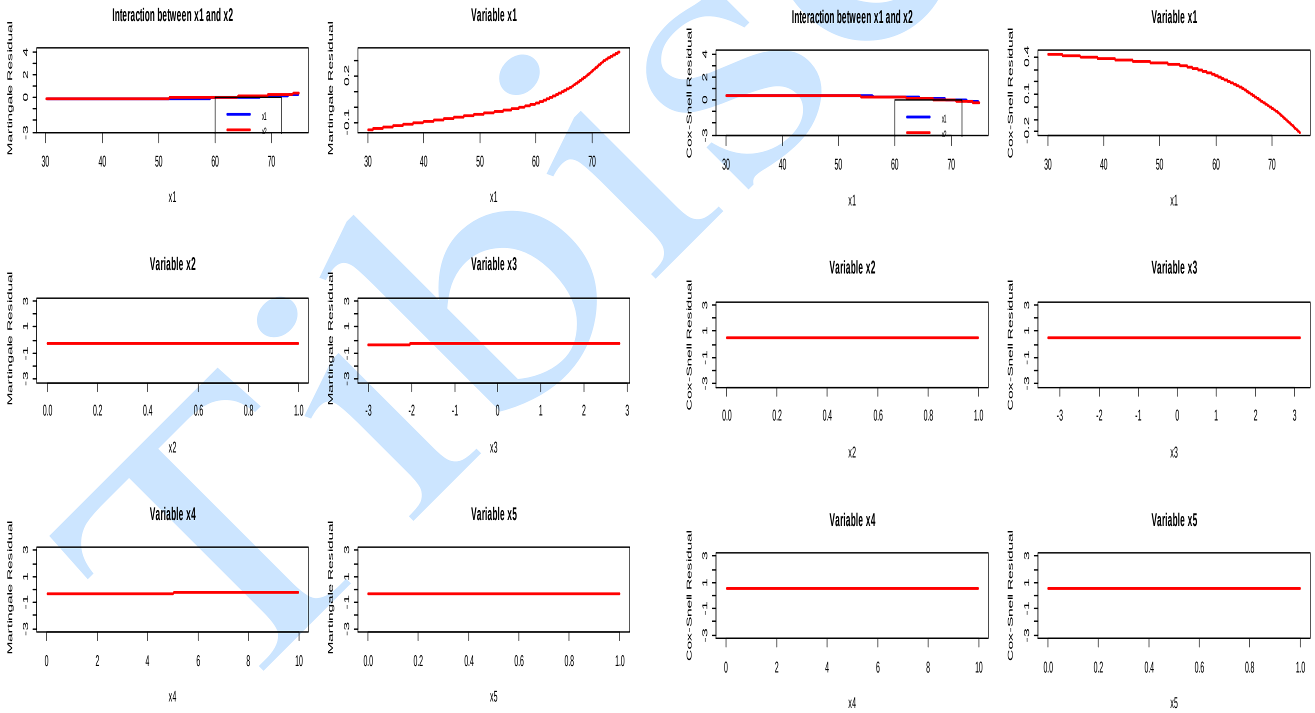

Fig. 8: Graphs of Survival MARS results under Weibull Survival times with penalized parameter $d = 5$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

Appendix C: Graphs of Survival MARS results for Simulated Gompertz Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters

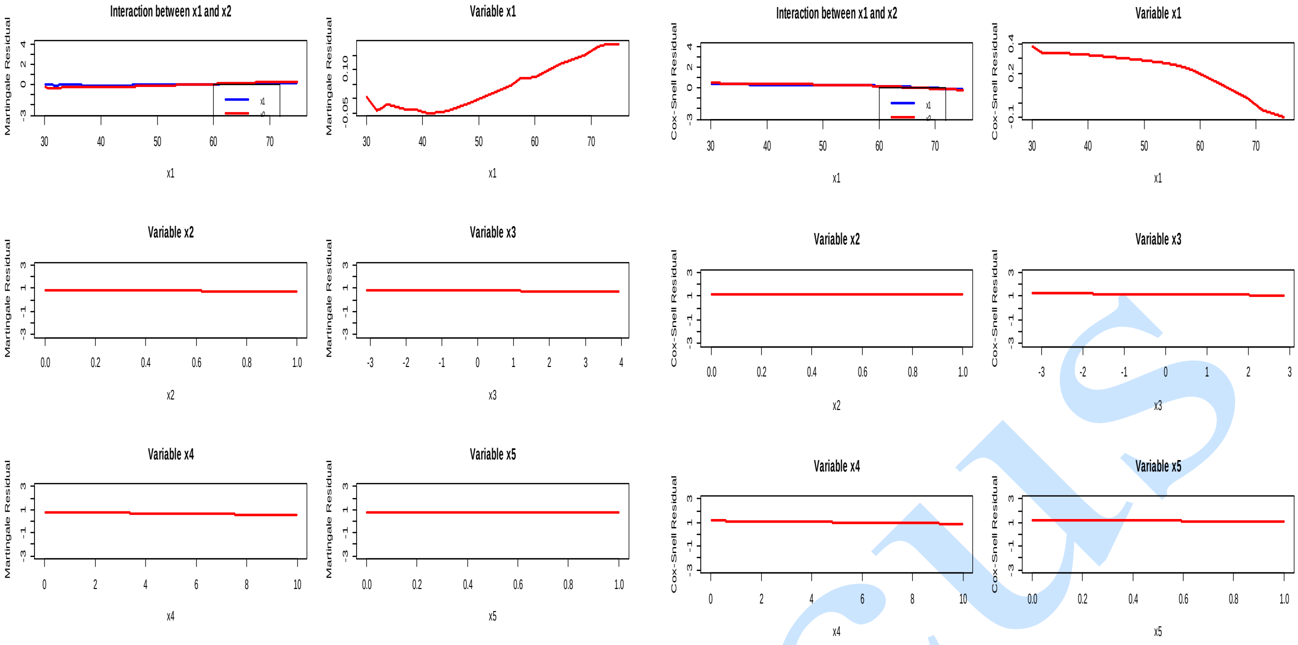

Fig. 9: Graphs of Survival MARS results under Gompertz Survival times with penalized parameter $d = 2$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

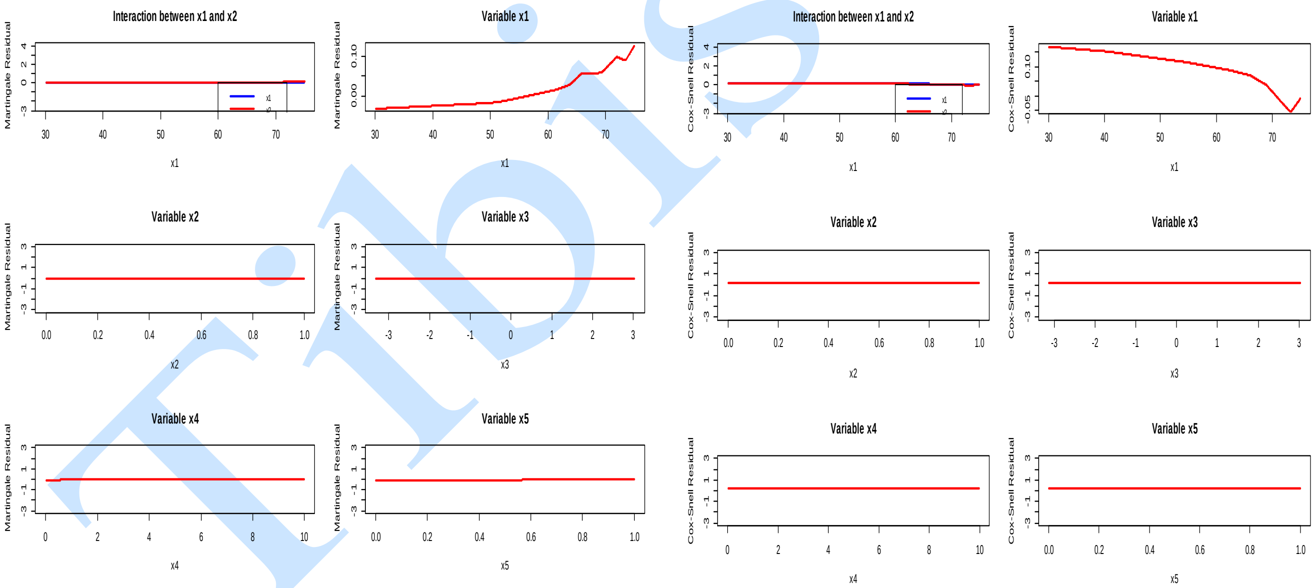

Fig. 10: Graphs of Survival MARS results under Gompertz Survival times with penalized parameter $d = 2$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

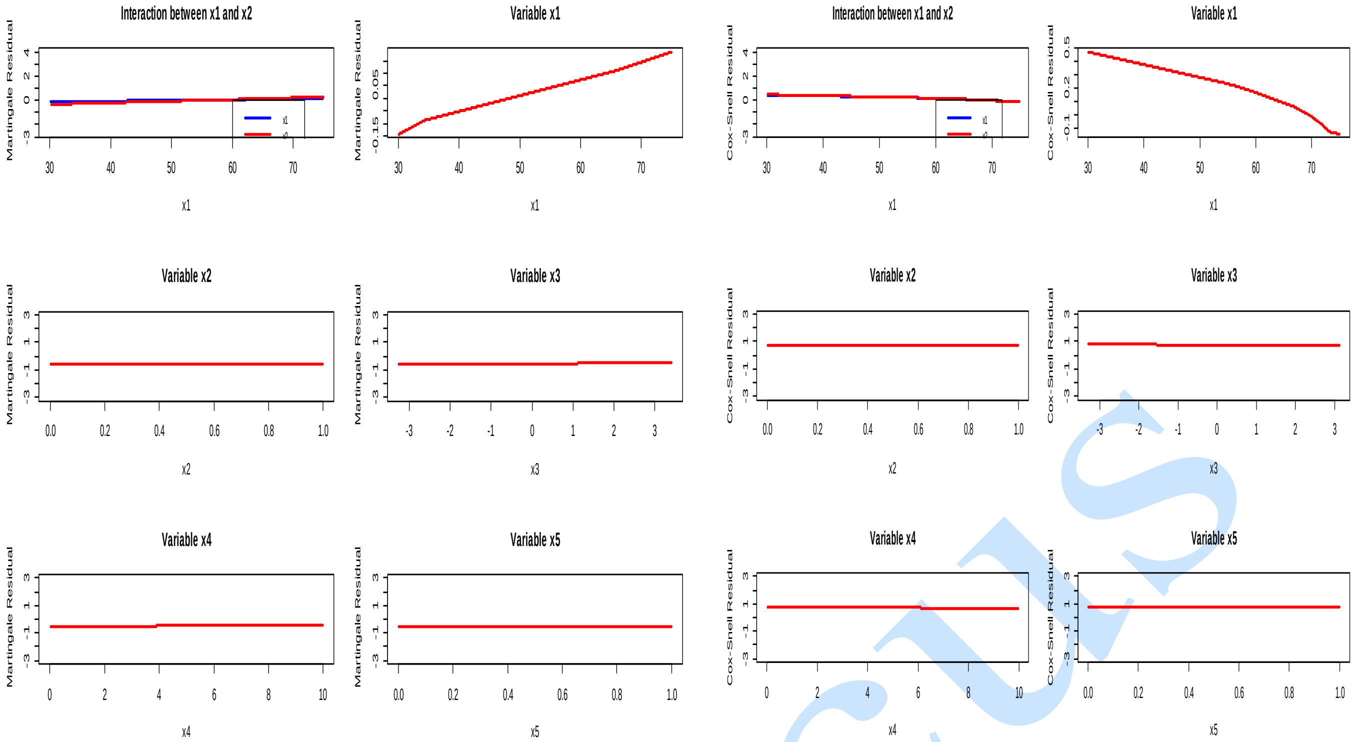

Fig. 11: Graphs of Survival MARS results under Gompertz Survival times with penalized parameter $d = 5$ at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals

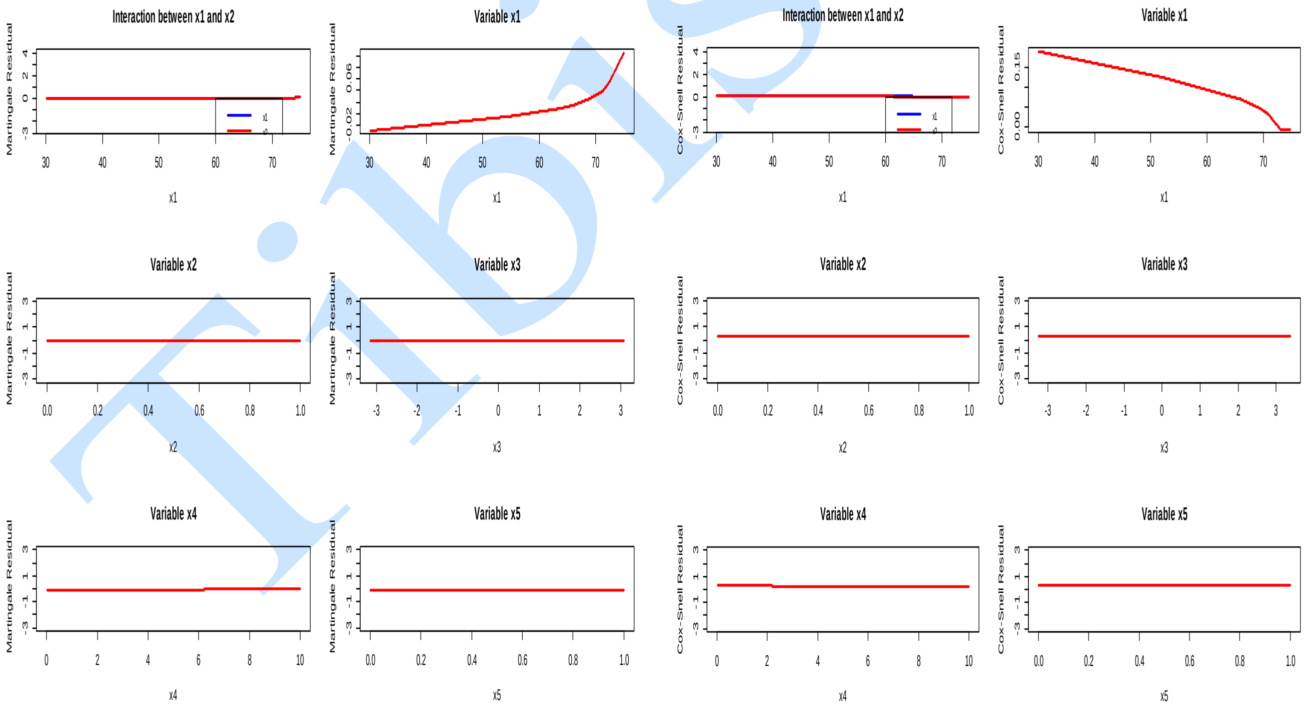

Fig. 12: Graphs of Survival MARS results under Gompertz Survival times with penalized parameter $d = 5$ at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals