Using Wavelet Shrinkage in the Cox Proportional Hazards Regression model (simulation study)

Prof. Dr. Taha Hussein Ali*   Jwana Rostam Qadir**
taha.ali@su.edu.krd   Jwana.Qadir@su.edu.krd

ABSTRACT

The proposed method in this paper dealt with the problem of data contamination in the Cox Proportional Hazards Regression model (CPHRM) by using Wavelet Shrinkage to de-noise data, calculating the discrete wavelet transformation coefficients for wavelets (Symlets and Daubechies), and thresholding methods (Universal, Minimax, and SURE), as well as thresholding rules (Soft and Hard). A software in the MATLAB language built for this propose will compare the proposed and classical method using simulation and real data. All the proposed methods have better efficiency than the classical method in estimating the Cox Proportional hazards model depending on both average of Akaike and Bayesian information criterion.

***Keywords:** Cox PH model, Wavelet Shrinkage, thresholding rules.*

إستخدام التقليص المويجي في أنموذج إنحدار كوكس للمخاطر النسبية (دراسة محاكاة)

تم في هذا البحث إقتراح معالجة مشكلة تلوث البيانات في أنموذج كوكس للمخاطر النسبية باستخدام التقليص المويجي، من خلال حساب معاملات التحويل المويجي المتقطع، للمويجات (Symlets)و (Daubechies) ، وطرائق العتبة (Universal) ، (Minimax) ، و (SURE) ، كذلك قواعد العتبة (الناعمة والصلبة). سيتم أيضاً المقارنة بين الطرائق المقترحة والتقليدية باستخدام المحاكاة والبيانات الحقيقية، التطبيق تم من خلال برنامج بلغة MATLAB والمصمم لهذا الغرض. كانت جميع الطرائق المقترحة تتمتع بكفاءة أفضل من الطريقة التقليدية في تقدير أنموذج كوكس للمخاطر النسبية اعتماداً على متوسط معيار معلومات (Akaike) و(Bayesian).

*Professor, Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University, Erbil.

1. Introduction

There are two major regression models used for censored data: proportional hazards (PH) model for Cox as a semi-parametric method (Cox, 1972) and accelerated failure time (AFT) models as a parametric method, e.g. Exponential, Weibull, and Lognormal distributions are parametric models lead to some benefits (Lawless, 1998). However, Cox regression is the most widely employed model in survival analysis.

In addition, the Cox model is widely used because it is reliable, the estimated risks are never negative, and the hazard ratio can be computed (Singh, 2011). The Cox model has played a vital role in applied survival analysis during the last three decades. The model and its software implementations have popularized survival analysis and made it accessible to researchers in varied disciplines who are not necessarily statisticians. It has been so successful that it is probably used in most practical analyses of the effects of covariates on survival (Royston, and Lambert, 2011). In order to calculate and test regression coefficients, even when all of the parametric model's assumptions are met, the CPHRM, has high efficiency when it is parametric models (for example, the Weibull and Gamma model with proportional hazards). When parametric model assumptions are not available (for example, when a Weibull and Gamma model are employed but the data is not from the Weibull and Gamma survival distribution respectively, resulting in an erroneous model choice), the CPHRM analysis is more efficient than parametric models (Harrell, 2015). Also, the CRPHM assumes two parts: that the proportional hazard (PH) is constant with time, while PH are variables have a log-linear relationship (Ekman, 2017).

On the other hand, wavelets are a good tool for the approximation of high dimensional functions, which feature dominant directions of the periodicity. One-dimensional shift invariant spaces and tensor-product wavelets are generalized to multivariate shift invariant spaces on non-tensor-product systems. The estimation of the non-parametric regression model in the (AFT) model under right random censorship and investigate the asymptotic rates of convergence of estimators based on thresholding of empirical wavelet coefficients (Linyuan et al. 2006). Wavelet Estimates with censored data are considered, to investigate the asymptotic rates of convergence of estimators by using thresholding of wavelet coefficients (Linyuan et al. 2007). Yogendra et al. (2010) discuss estimation of the

density derivative by using wavelets methods by using randomly-censored data and extend the results to asymptotic convergence rates due to Prakasa Rao (1996) and Chaubey et al. (2008) under random censorship model. Eddy (2011) suggest approach in estimating the function suitable compactly supported wavelets like the Daubechies, Symlets or Coiflets family of wavelets, the smoothness and time-frequency properties of these wavelets allow us to find an asymptotically estimators of the slope coefficient of the linear model. Rogério (2016) suggest extraction of an observation in the presence of random noise by wavelet shrinkage has been studied under assumptions that the contaminate is independent and identically distributed and that the samples are evenly spaced with time. Xing et al. (2017), discuss the estimation a models with censored data by using wavelet method when the survival function and the censoring times has a stationary α-mixing sequence, and of the wavelet estimators for varying functions. Christophe et al. (2019) suggest a deal with the estimation of a non-parametric regression with both additive and contaminate, for uniform multiplicative contaminate is considered, and develop a projection estimator by using a several wavelets. Jinru et al. (2020), explained the wavelet estimators of censored mixture density and discuss their point wise asymptotic convergence rates.

2. Cox proportional hazard Regression model:

The CPHRM is as in the following formula:

$$h(x,t) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i x_i\right] \tag{1}$$

Where $x_1, x_2, \cdots, x_p$ is a collection of covariate $h_0(t)$ for all explanatory factors, the baseline hazard time t represents the hazard for a person with a value of 0. $\beta_0, \beta_1, \cdots, \beta_p$ are regression coefficient which is estimated by the partial likelihood method (Aako and Are, 2020).

The main advantages of the Cox PH model

1- without estimate $h_0(t)$ we can estimate the parameters (β_i)
2- we don't have to assume that $h_0(t)$ follows a Weibull model, or a Gamma model, or any another parametric model.

   CPHRM assumptions:

1- The $h_0(t)$ is non-parametric.
2- On the log-rate scale, covariate effects are additive and linear.
3- Proportional hazards: Over time, the ratio of hazard rates for two groups remains constant.

4- Time t is "automatically" adjusted.

Time-dependent and time-independent variables are the two types of covariates used in survival analysis.

2.1. Tim-dependent Covariates

Whereas time-dependent covariates are those whose values do change over time. Time-dependent covariates are further categorized into two types, internal and external covariates. Kalbfleisch and Prentice claim that (Aako and Are, 2020), If a time-dependent covariate meets the criteria, it is termed external.

$$P[X(t)|X(u),T > u] = P[X(t)|X(u),T = u]$$

for all *u, t,* such that $0 < u < t$. This means that at although the covariate could affect the hazard function over time, its upcoming path up to every time $t > u$ The occurrence of a failure at time u has no effect. in another way, exogenous variables do not need the survival of a subject to exist. An external covariate is one whose value is known in advance at any moment in the future, such as a subject's age or a drug's recommended dose during a research (Liu, 2005).

In the other hand Time-dependent variables may be readily incorporated into the model to account for characteristics that vary over time. the hazard function is defined as:

Let $z_{ij}(t)$ are independent variables (j^{th} covariate of the ith unit under observation), for $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, p$, and t is an observation with time scale. The notation $Z_{ij}(t)$ indicates that the value of $z_{ij}(t)$ varies as a values with time scale. Then the CPHRM with time-dependent covariates specifies that the hazard rate for the ith individual as in the following formula:

$$h_i(t) = h_0(t)\exp(Z_i(t)\beta) \qquad (2)$$

$h_0(t)$ is the baseline hazard rate, $Z_{ij}(t)$ is a vector by $(1 \times p)$ dimension of independent variables for unit i and that may be either fixed or dependent time, and β is a regression coefficients for vector by $(1 \times p)$ dimension. The advantages of the CPHRM over other types of time-to-event methods is the $h_0(t)$ can be left unspecified in practice. The functional form that a practitioner should perform is that $h_0(t)$ is a non-negative value of t. For researchers with weak substantive theory for the hazard shape when $Z_{ij}(t) = 0$, The CPHRM model is more flexible. However, because it assumes proportional hazards (PHs), the CPHRM places a significant

limitation on the data. Time-dependent Value of variable differs over time Hazard ratio (HR) formula (3), (Gail et al., 2007):

$$\widehat{HR} = exp\left[\sum_{i=1}^{p} \hat{\beta}_i \left(X_i^* - X_i\right)\right] \tag{3}$$

When $X_i^* = (X_1^*, X_2^*, \dots, X_p^*)$ and $X_i = (X_1, X_2, \dots, X_p)$

2.2. Time-independent Covariates

Independent variables (Covariates) whose values do not vary with time are said to be time-independent (Aako and Are, 2020). Time-independent The value of the variable remains constant throughout time, whereas the exponential expression includes X but not t. The X in this case are known as time-independent X. Moreover, the hazard ratio comparing any two specifications of X predictors is constant throughout time, according to the (PH) assumption underpinning the Cox PH model. This indicates that the risk posed by one individual is proportionate to the risk posed by any other individual, with the proportionality constant remaining constant throughout time. the cox PH model with time-independent covariates in the formula (4), (Gail et al., 2007).

$$h(X_i, t) = h_0(t) \exp\left[\sum_{j=1}^{p} x_{ij} \beta_j\right] \tag{4}$$

2.3. Time-independent and dependent covariates

Both time-independent and time-dependent predictor variables, we can write the extended Cox model that incorporates both types, as in the following formula:

$$h(X_i, t) = h_0(t) \exp\left[\sum_{j=1}^{p_1} x_{ij} \beta_j + \sum_{r=1}^{p_2} x_{ir}(t) \beta_r\right] \tag{5}$$

Where $X_i = X_1, X_2, \cdots, X_{p_1}$ time-independent, $X_r(t) = X_1(t), X_2(t), \cdots, X_{p_2}(t)$ time-dependent.

3. Wavelet Shrinkage

Wavelet shrinkage is well established technique for removing the noise present in the observation, while preserving the significant features of the original data (Donoho, 1994). The wavelet shrinkage based on thresholding of the wavelet coefficients.

3.1. Wavelet

Wavelet are small waves that can be grouped together to form larger waves or different waves. A few fundamental waves were used, stretched in infinitely many ways, and moved in infinitely many ways to produce a wavelet system that could make an accurate model of any wave. Consider generating an orthogonal wavelet basis for functions $f \in L^2(R)$ (the space of square integrable real functions), starting with two parent wavelet: the scaling function ϕ (also called farther wavelet) and the mother wavelet ψ. Other wavelets are then generated by dilations and translations of ϕ and ψ (Donald et al., 2004). The dilation and translated of the functions are defined by formulas (6) and (7).

$$\phi_{k,q}(y) = 2^{k/2}\phi(2^k y - q) \qquad\qquad k,q \in z \qquad (6)$$

$$\psi_{k,q}(y) = 2^{k/2}\psi(2^k y - q) \qquad\qquad k,q \in z \qquad (7)$$

The discrete wavelet transform (DWT) is a widely applicable observation processing algorithm that is used in various applications, for instance, science, engineering, mathematics and computer science. DWT decomposes an observation by using scaled and shifted versions of a compact supported basis function (mother wavelet), and provides multiresolution representation of the observation (Iolanda, 2007). Given a vector of an observation $\mathbf{y}$ consisting of 2^k observations, where k is an integer and the DWT of $\mathbf{y}$ due to formula (8).

$$W = w\mathbf{y} \qquad (8)$$

Where w is wavelet matrix with $(n \times n)$ dimensions, W is a vector with $(n \times 1)$ dimensions including both scaling and wavelet coefficients. The vector of wavelet coefficients can by organized into $(k+1)$ vectors. $W = [W_1, W_2, ..., W_k, V_{k0}]^T$. At each DWT, the approximation coefficients are divided into bands using the same wavelet as before, with the result that the details are appended with the details of the latest decomposition, as in the following formula:

$$\mathbf{y} = Ww^T = \sum_{k=1}^{k_0} W_k^T W_k + V_{k_0}^T V_{k_0} \qquad (9)$$

At each level (k), the observations can be reconstructed of the de-noise data (reduce of the contamination) by the inverse DWT (Ramazan et al., 2002).

3.2 Thresholding

Thresholding is the simplest method of non-linear wavelet de-noising, in which sub dividing the wavelet coefficient in to two sets, one of which represents signal while the other represents noise. There are different rules to apply the thresholds of the wavelet coefficients, and several different methods for choosing a threshold value exist such as:

A. Universal Threshold Method

Donoho and Johnstone (1994) submitted universal threshold method, which is given by formula (10).

$$\delta^U = \hat{\sigma}_{MAD} \sqrt{2 Log(n)} \tag{10}$$

Where $\hat{\sigma}_{MAD}$ is the standard deviation estimator of details coefficients, and equal to $MAD/0.6745$. Where MAD is the median absolute deviation of the wavelet coefficients at the finest scale.

B. Minimax Threshold Method

The optimal minimax threshold method submitted by Donoho and Johonstone, (1994) as an improvement to the universal threshold method, Minimax is based on an estimator $\tilde{f}$ that attains to the minimax risk, as:

$$\tilde{R}(F) = \inf_{\tilde{f}} \sup_{f \in \tilde{R}(F)} R(\tilde{f}, f) \tag{11}$$

Where

$$R(\tilde{f}, f) = \frac{1}{n} \sum_{i=1}^{n} E[\tilde{f}, f]^2 \tag{12}$$

Where $f = f(x_i)$ and $\tilde{f} = \tilde{f}(x_i)$, denote the vectors of true and estimated sample values. The threshold minimax estimator is different from universal counter parts, in which the minimax threshold method is concentration on reducing the overall mean square error (MSE) but the estimates are not over-smoothing.

C. SURE Threshold Method

The sure threshold proposed by Donoho and Johonstone (1994), which based upon the minimization of stein's risk estimator. In sure threshold method specifies a threshold estimate of δ_k at each level k for the wavelet coefficients, and then for the soft threshold estimator we have.

$$SURE(\delta, W) = n - 2 \neq \{k : |W_k| \leq \delta\} - \sum_{k=0}^{d} \min(|W_k|, \delta) \tag{13}$$

Where $[W_k : k = 1, 2, \cdots, d]$ be a wavelet coefficients in the kth level, and then, select δ^s that minimizes SURE (δ, W).

$$\delta^S = \arg\min SURE(\delta, W) \tag{14}$$

3.3 Thresholding Rules

There are many rules for the thresholding. The two types used in this research will be discussed.

A. Soft Thresholding

The other standard technique for wavelet de-noising is Soft thresholding of the wavelet coefficient, also proposed by Donoho and Johnostone, which is defined as follows (Jeena, 2013).

$$Wn^{(s)} = sign\{Wn\}\big(|Wn| - \delta\big)_+ \tag{15}$$

Where

$$Sign\{Wn\} = \begin{bmatrix} +1 & if & Wn > 0 \\ 0 & if & Wn = 0 \\ -1 & if & Wn < 0 \end{bmatrix} \tag{16}$$

and

$$\big(|Wn| - \delta\big)_+ = \begin{bmatrix} \big(|Wn| - \delta\big) & if & \big(|Wn| - \delta\big) \geq 0 \\ 0 & if & \big(|Wn| - \delta\big) < 0 \end{bmatrix} \tag{17}$$

Coefficients smaller than threshold are set to zero, and additionally all coefficients greater than threshold are reduced by the amount of threshold. Thus, the Soft thresholding is a continuous mapping.

B. Hard Thresholding

Donoho and Johnstone proposed Hard thresholding, it is a simplest scheme thresholding interpreting the statement of (keep or kill). The Hard thresholding used straightforward technique for implementing wavelet de-noising (Katsuyuki, 2021). The wavelet coefficient is set to the vector $Wn^{(H)}$ with element.

$$Wn^{(H)} = \begin{bmatrix} 0 & if & |Wn| \leq \delta \\ Wn & if & |Wn| > \delta \end{bmatrix} \tag{18}$$

coefficients exceeding δ are left untouched, while smaller than or equal to δ are eliminated or set to 0. Thus, the operation of hard thresholding is not continuous mapping.

4. Proposed method

The proposed method included dealt with the problem of Cox Proportional hazards model $h(t)$ data contamination, by using Wavelet Shrinkage. First, the discrete transformation coefficients $W(t)$ for a wavelet $w(t)$ (e.g. Symlets and Daubechies wavelets) composed of two parts (wavelet and scaling functions) are calculated from formula (19):

$$W(t) = w(t) \times h(\mathrm{t}); \quad t = 1, 2, \cdots, n \tag{19}$$

Using the first level of discrete wavelet coefficients $W_{1,0}$, $W_{1,1}$, ... , $W_{1,(n/2)-1}$ The threshold level δ is estimated by one of the methods (e.g. SURE, Minimax, and Universal threshold) for estimating the threshold level as formulas (10), (11), and (14).

Thresholding rules, Soft and Hard are used to keep or kill the discrete wavelet coefficients obtained from the formulas (15) and (18), depending on the threshold level estimated δ, such that discrete wavelet coefficients below of δ are zeroed (kill) and above of δ are keep. More clearly, large coefficients that are greater than δ remain unchanged, while those that are less than or equal to δ are deleted or are a set of zero. Thus, we get the modified discrete wavelet transformation coefficients $MW(t)$, then it is used to compute the inverse of the modified discrete wavelet transform as in formula (20).

$$hw(\mathrm{t}) = Inv(MW(t)) \tag{20}$$

Finally, the proposed wavelet Cox Proportional hazards model is obtained in the formula (21) and which has less contamination.

$$hw(x,t) = h_0(t) \exp\left[\sum_{i=1}^{p} \beta_i x_i\right] \tag{21}$$

Formula (21) represents the proposed model without the time-dependent covariates, while formula (22) represents the proposed model with the time-dependent covariates and which also has less contamination.

$$hw(x,t) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{k=1}^{p_2} \beta_k x_k(t)\right] \tag{22}$$

5. Evaluation criteria

Akaike information criterion (AIC) and Bayesian information criterion (BIC) which depend on Log-likelihood (LL) will be used as selection criteria for the models. The model with the lowest value of AIC

and BIC term appears the best model to Cox PH regression (Rinku and Manash 2016).

$$AIC = -2(LL) + 2K/p \tag{23}$$

$$BIC = -2(LL) + K.\text{Ln}(n) \tag{24}$$

6. Experimental and Application

To compare between the classical and the proposed method in terms of efficiency and accuracy of the estimated model, an experimental part was done by simulating the Cox-proportional hazards (PH) model, then an applied part of the real data based on AIC and BIC. And by designing a program in MATLAB (version 2020a) dedicated to this purpose (Appendix).

6.1 Experimental Part

For case the time-independent and dependent covariates, simulation data visualizations (Appendix -program-1). Three cases were selected for the sample size (100, 200, and 300). It was assumed that there are five covariates, three of which are not time-dependent and two are time-dependent from an autoregressive model AR(1), with $\rho_1 = 0.8$ and $\rho_2 = -0.8$, for the first simulation with n = 100. The vector regression parameters were also imposed $[0.5 \quad 0.75 \quad 1 \quad 0.5 \quad -0.5]^T$. Noises with a Laplace distribution $L(0,0.25)$ are added to the Cox PH model, dependent variable without noise and with noise for the first simulation with n = 100 shown figure (1). The 60% data is censored, and 40% is uncensored, with constant hazard rate (0.1) as initial value.
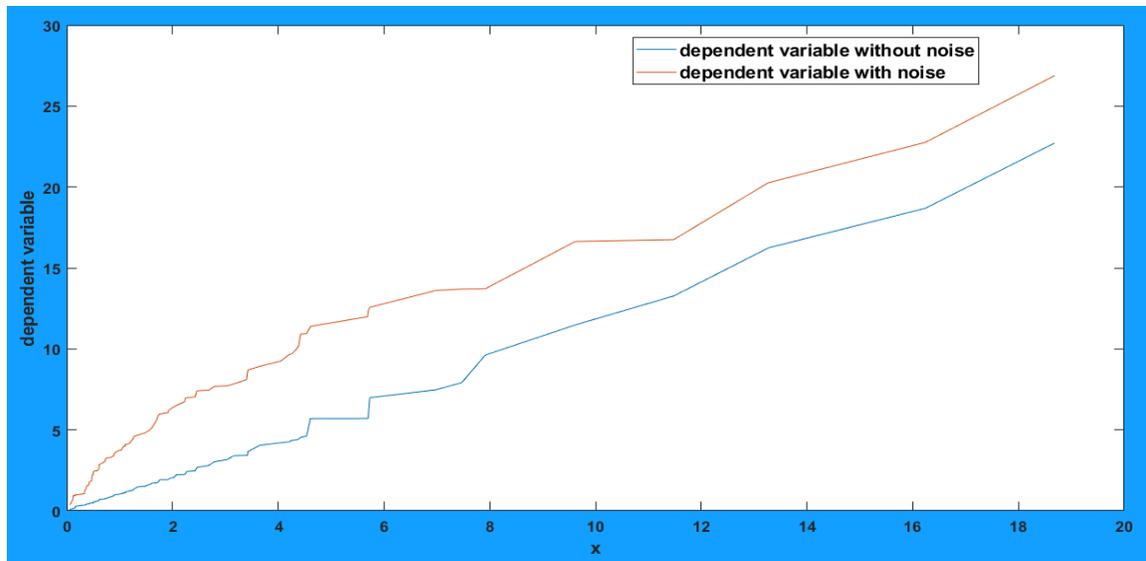
Figure (1): Dependent variable without noise and with noise

For the purpose of the comparison between the proposed and classical method in estimating the CPHRM, also the experiment was repeated to (1000) times and the average criteria for AIC and BIC was calculated. Two wavelets (Sym2) and (db13) were used with different methods in estimating the threshold level (SURE, Minimax, and Universal), for two threshold rule (Soft and Hard), and for different samples (100, 200, and 300). The results are summarized in Tables (1-3).

Table (1): Average of criteria AIC and BIC for (1000) when $n = 100$

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	293.69	304.73
	Sym2	SURE	Hard	294.06	305.09
	Sym2	Minimax	Soft	293.41	304.43
	Sym2	Minimax	Hard	294.16	305.19
	Sym2	Universal	Soft	293.07	304.09
	Sym2	Universal	Hard	293.30	304.33
	db13	SURE	Soft	293.54	304.57
	db13	SURE	Hard	294.18	305.21
	db13	Minimax	Soft	293.29	304.31
	db13	Minimax	Hard	294.09	305.12
	db13	**Universal**	**Soft**	**293.04**	**304.07**
	db13	Universal	Hard	293.49	304.51
Classical				294.35	305.37

Table (2): Average of criteria AIC and BIC for (1000) when $n = 200$

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	692.56	707.05
	Sym2	SURE	Hard	693.36	707.85
	Sym2	Minimax	Soft	691.68	706.17
	Sym2	Minimax	Hard	693.17	707.66

	Sym2	Universal	Soft	689.92	704.42
	Sym2	Universal	Hard	691.06	705.55
	db13	SURE	Soft	691.21	705.70
	db13	SURE	Hard	692.79	707.29
	db13	Minimax	Soft	690.15	704.64
	db13	Minimax	Hard	692.97	707.47
	db13	**Universal**	**Soft**	**688.08**	**702.57**
	db13	Universal	Hard	689.27	703.76
Classical				693.09	707.58

Table (3): Average of criteria AIC and BIC for (1000) when $n = 300$

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	1133.2	1149.7
	Sym2	SURE	Hard	1134.1	1150.7
	Sym2	Minimax	Soft	1132.4	1148.9
	Sym2	Minimax	Hard	1134.3	1150.8
	Sym2	Universal	Soft	1130.6	1147.1
	Sym2	Universal	Hard	1131.4	1147.9
	db13	SURE	Soft	1131.2	1147.7
	db13	SURE	Hard	1133.5	1150.0
	db13	Minimax	Soft	1130.5	1147.0
	db13	Minimax	Hard	1134.0	1150.5
	db13	**Universal**	**Soft**	**1128.0**	**1144.5**
	db13	Universal	Hard	1129.8	1146.3
Classical				1134.4	1150.9

Tables (1–3) show that all the proposed methods have better efficiency than the classical method in estimating the Cox PH model depending on both average of criteria (AIC and BIC) for various selected samples, except the case ($n = 200$), for the (Sym2) wavelet, SURE and Minimax threshold method, Hard rule. Also, (db13) wavelet with Universal threshold method and Soft threshold rule was the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of both criterions and for various selected samples (AIC = 293.04, BIC = 304.07), (AIC = 688.08, BIC = 702.57), and (AIC = 1128.0, BIC = 1144.5) respectively. For most simulation experiments, (db13) Wavelet was better than (Sym2), Also Universal threshold method was better than SURE and Minimax, and Soft rule better than Hard for all cases. Note that the average values of the two criterions increase with the increase in the sample size, any decrease in the efficiency of the proposed and classical estimated models if the sample size increases.

6.2 Application Part

This Application shows how to fit of the CPHRM from panel data, **years of observed** of loan status represent dependent variable. For the model includes only time independent predictors, any information that remains constant throughout the life of the loan. Just a **set of points** and **vintage information**, when creating loans as an independent predictor of time, because it is the degree given to borrowers at the beginning of the loan, and the return is constant throughout the life of the loan.

CPHRM is a semiparametric method to adjusting survival rate estimates to quantify the impact of independent variables. The method represents the effects of independent variables as a multiplier of a $h_0(t)$. The hazard function is the nonparametric part of the Cox PH regression function, whereas the effect of the independent variables is a log-linear regression. To fit the model, the sample data is randomly split into two parts. First, split the data into training (60% of data equal to 58092 observations from 96820). The hypotheses to be tested are as follows:

H_0 : The model is unfit vs. H_1 : The model is fit

$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0 \quad \forall \quad j = 1,2$

Table (4): Classical Cox PH Model

Cases available		beta	S.E.	Z	p-value	Chi-square	p-value	AIC	BIC
Event	3917	-0.6960	0.0368	-18.888	0.000	1017.5	0.000	84470	84493
Censored	54175	-1.2747	0.0454	-28.060	0.000				
Total	58092								

Table (4) shows that the data included (3917) observations event and (54175) censored, and the classical Cox PH Model is fit, because the value of chi-square (1017.5) for overall (score) is greater than its tabulated value under the significance level ($\alpha = 0.05$) and degrees of freedom (2) which is equal to (5.99), p-value equal to zero and its less than α. The classical Cox regression coefficients (-0.6960 and -1.2747) are significant because the absolute values of Z (18.888 and 28.060) respectively are greater than tabulated value (1.96), p-values equal to zero and its less than α.

The efficiency of the classical CPHRM is represented by the criterion AIC, which is equal to (84470), and the criterion BIC, which is equal to (84492). The baseline cumulative *HR* can be converted to the hazard rate *h*, except for adding a step for analysis. The Classical CPHRM assumes that the observation time is measured as a continuous variable. The **coxphfit** function in MATLAB supports methods for handling joins in a time variable.

Also, CPHRM will be estimated by proposed method, depending on the wavelet shrinkage represented by the (sym2) wavelet, with SURE threshold method for estimating the threshold level, and using the soft threshold. Using the same covariates (score group and vintage information) and data generated previously. To fit the proposed CPHRM, and to test the previous hypotheses.

Table (5): Proposed Cox PH Model

Cases available		beta	S.E.	Z	p-value	Chi-square	p-value	AIC	BIC
Event	3917	-0.7037	0.0368	-19.098	0.000	1038.7	0.000	83754	83776
Censored	54174	-1.2861	0.0454	-28.311	0.000				
Censoredb	1								
Total	58092								

Table (5) shows that the data included (3917) observations event, (54174) censored, and censored cases before the earliest event in a stratum (Censoredb) equal to one. The proposed Cox PH Model is fit, because the value of chi-square (1038.7) for overall (score) is greater than its tabulated value under the significance level α and degrees of freedom (2) which is equal to (5.99), p-value equal to zero and its less than α. The proposed Cox regression coefficients (-0.7037 and -1.2861) are significant because the absolute values of Z (19.098 and 28.311) respectively are greater than tabulated value (1.96), p-values equal to zero and its less than α. The efficiency of the proposed Cox PH Model is represented by the criterion AIC, which is equal to (83754), and the criterion BIC, which is equal to (83776). The baseline cumulative hazard rate H can be converted to the hazard rate h as before, the proposed Cox PH model assumes that the observation time is measured as a continuous variable and after the wavelet shrinkage procedure, the data became continuous, which was used in the account classical Survival, One Minus Survival, Hazard, and LML Function at mean of covariates.

To comparison of the proposed method (Wavelet shrinkage) with the classical method of estimating the Cox PH model. Starting with the data of the first sample, table (4) and (5) shows that the proposed method is better than the classical method for the data of the first experiment from the simulation, because the values of AIC and BIC (83754 and 83776) respectively for the proposed method was less than AIC and BIC (84470 and 84492) respectively for the classical method. The Chi-square value to test the significance of the proposed estimated model was greater than the

classical model, and also its estimated parameters with the stability of the standard error values for both models.

The proposed Cox PH model estimated has a continuous variable, while the classical method for the data of the first experiment from simulation had a discrete variable, as clear in computing and plotting of proposed and classical functions for Survival, One Minus Survival, Hazard, and LML Function at mean of covariates, in the Figures (2-5).
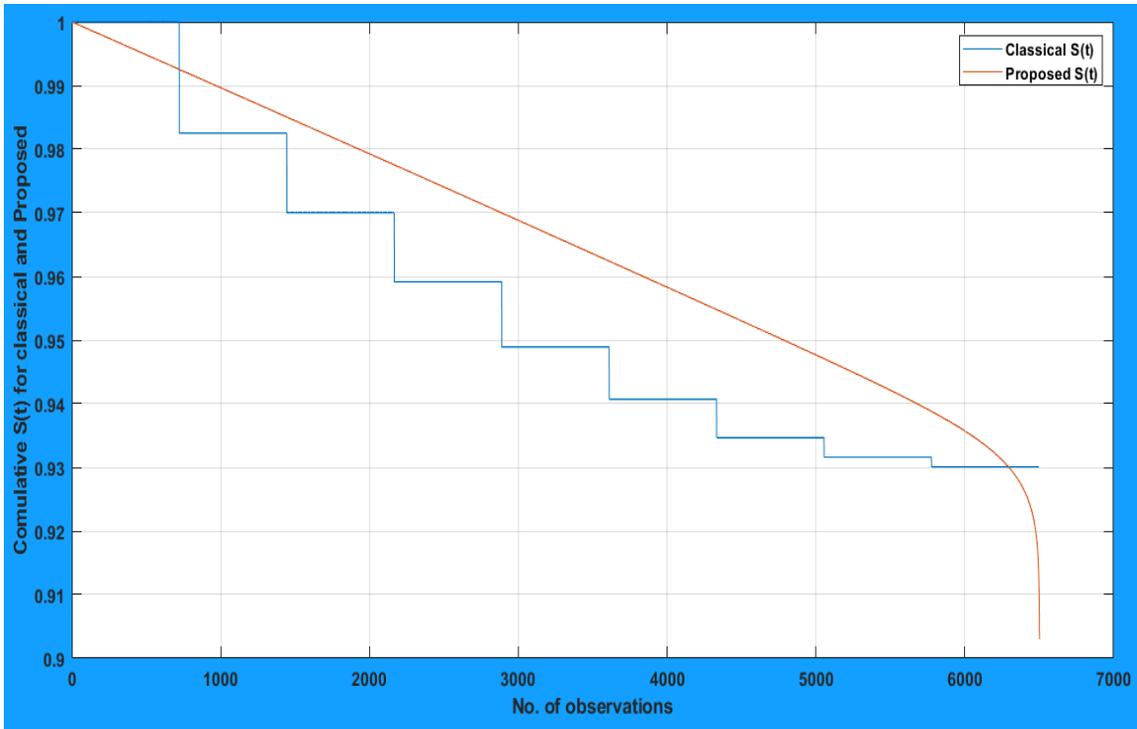

Figure (2): Classical and proposed Survival Function at mean of covariates
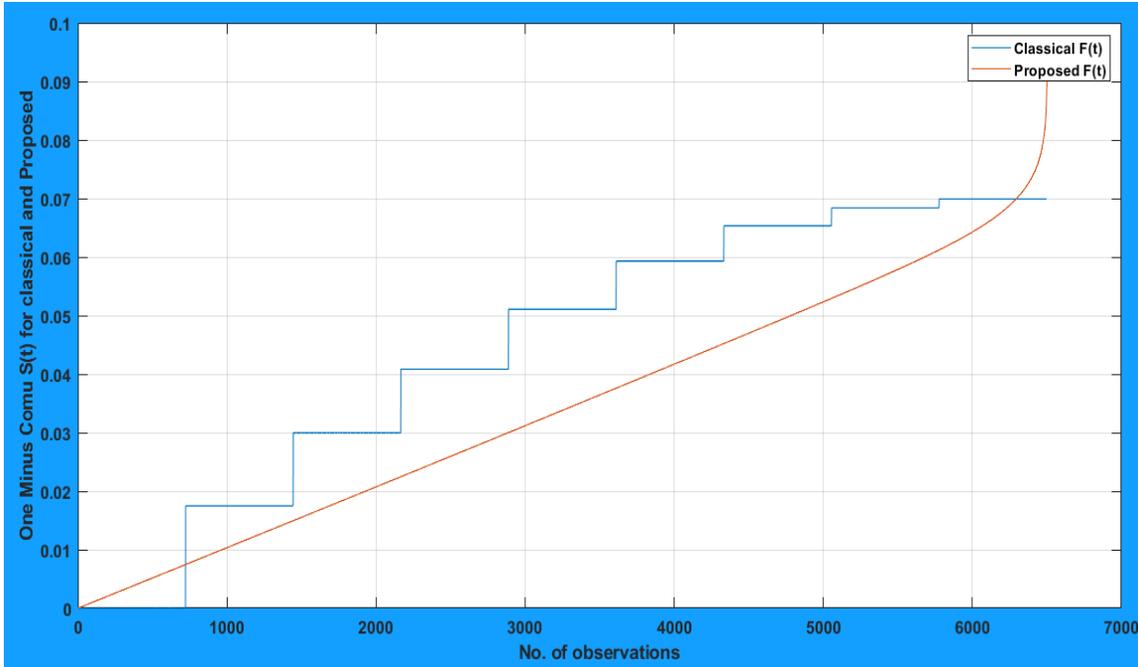
Figure (3): **Classical and Proposed One Minus Survival Function at mean of covariates**
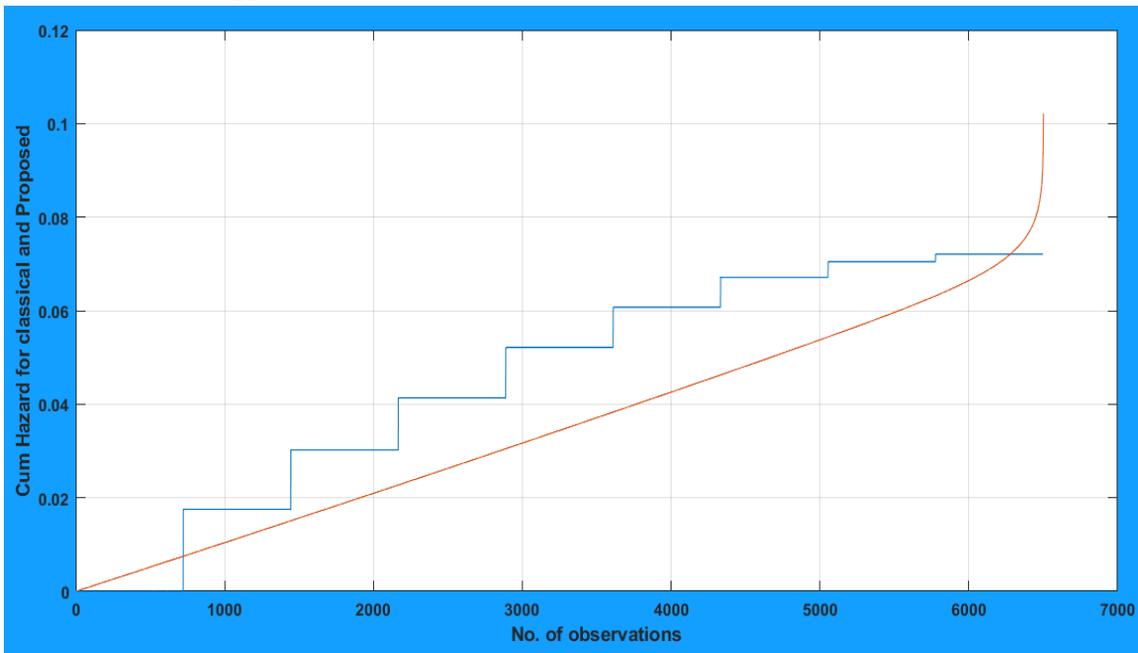

Figure (4): Classical and Proposed Cum Hazard Function at mean of covariates
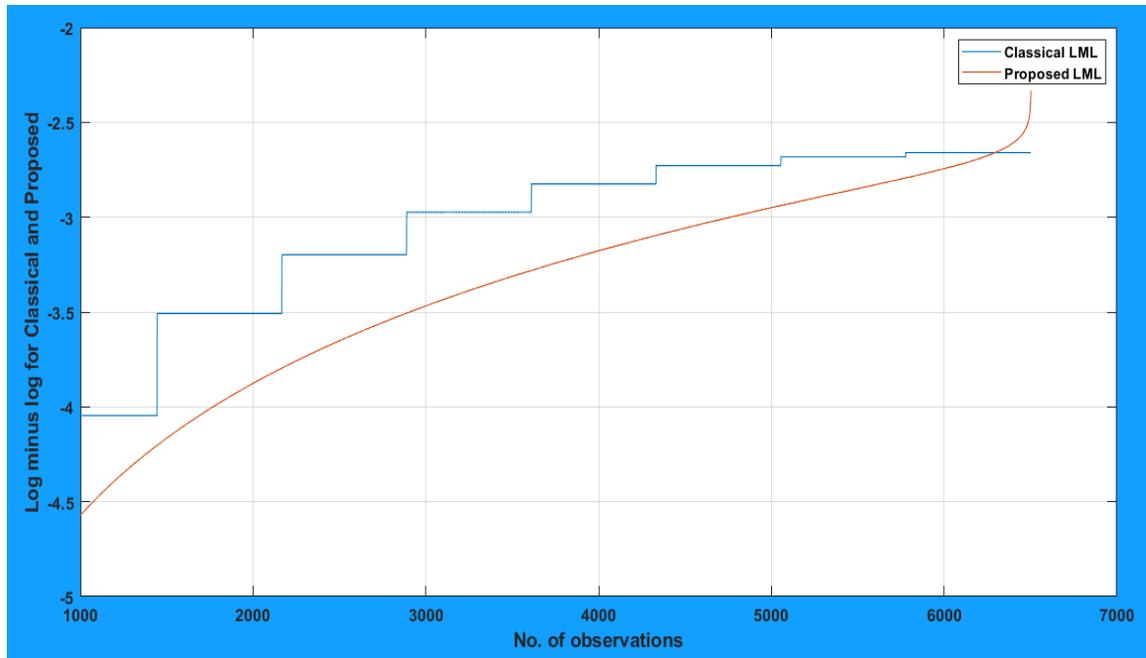
Figure (5): Classical and Proposed LML Function at mean of covariates

For the purpose of generalizing the results of the comparison between the proposed and classical method in estimating the Cox PH model, the application was repeated to (1000) times and the average criteria for AIC and BIC was calculated. Two wavelets (Sym2) and (db13) were used with different methods in estimating the threshold level (SURE, Minimax, and Universal), for two threshold rule (Soft and Hard), and for different samples (0.9 of the original data set equal to 87138, 0.6 of the original data set equal to 58092, and 0.3 of the original data set equal to 29046). The results are summarized in Tables (6-8).

Table (6): Average of criteria AIC and BIC for (1000) times, when $n = 87138$

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	129800	129820
	Sym2	SURE	Hard	130090	130120
	Sym2	Minimax	Soft	125250	125270
	Sym2	Minimax	Hard	125610	125630
	Sym2	Universal	Soft	121250	121280
	Sym2	Universal	Hard	121740	121760
	db13	SURE	Soft	129740	129760
	db13	SURE	Hard	130007	130100
	db13	Minimax	Soft	123450	123480
	db13	Minimax	Hard	124010	124030
	db13	**Universal**	**Soft**	**120610**	**120640**
	db13	Universal	Hard	120790	120820
Classical				130910	130930

Table (7): Average of criteria AIC and BIC for (1000) times, when $n = 58092$

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	83367	83389
	Sym2	SURE	Hard	83562	83584
	Sym2	Minimax	Soft	80331	80353
	Sym2	Minimax	Hard	80572	80594
	Sym2	Universal	Soft	77667	77689
	Sym2	Universal	Hard	77992	78014
	db13	SURE	Soft	83324	83346
	db13	SURE	Hard	83546	83568
	db13	Minimax	Soft	79133	79155
	db13	Minimax	Hard	79503	79525
	db13	**Universal**	**Soft**	**77239**	**77261**
	db13	Universal	Hard	77359	77381
Classical				84103	84124

Table (8): **Average of criteria AIC and BIC for (1000) times, when $n = 29046$**

Method	Wavelet	Threshold Method	Threshold Rule	AIC	BIC
Proposed	Sym2	SURE	Soft	38984	39004
	Sym2	SURE	Hard	39082	39102
	Sym2	Minimax	Soft	37466	37487
	Sym2	Minimax	Hard	37586	37607
	Sym2	Universal	Soft	36135	36156
	Sym2	Universal	Hard	36299	36320
	db13	SURE	Soft	38964	38984
	db13	SURE	Hard	39075	39095
	db13	Minimax	Soft	36867	36888
	db13	Minimax	Hard	37052	37072
	db13	**Universal**	**Soft**	**35921**	**35941**
	db13	Universal	Hard	35981	36002
Classical				39353	39374

Tables (6–8) show that all the proposed methods have better efficiency than the classical method in estimating the Cox PH model depending on both average of criteria (AIC and BIC) for various selected samples. And db13 wavelet with Universal threshold method and Soft threshold rule was the best efficient compared with all other proposed methods and with the classical method because it has the lowest average of both criterions and for various selected samples (AIC = 120610, BIC = 120640), (AIC = 77239, BIC = 77261), and (AIC = 35921, BIC = 35941) respectively. For all applications, (db13) Wavelet was better than (Sym2), Universal threshold method was better than SURE and Minimax, and Soft rule better than Hard. Also note that the average values of the two criterions increase with the increase in the sample size, any decrease in the efficiency of the proposed and classical estimated models if the sample size increases.

7. Conclusions

1. The proposed methods (Wavelet shrinkage) have better efficiency than the classical method in estimating the Cox PH model depending on both average of criteria (AIC and BIC) for various selected samples (for simulation and application).

2. Db13 wavelet with Universal threshold method and Soft threshold rule was the best efficient compared with all other proposed methods and with the classical method for various selected samples (for simulation and applications).

3. For most simulation experiments and applications, (db13) Wavelet was better than (Sym2), Also Universal threshold method was better than SURE and Minimax, and Soft rule better than Hard for all cases.

4. The average values of the two criterions increase with the increase in the sample size, any decrease in the efficiency of the proposed and classical estimated models if the sample size increases (for simulation and applications).

5. For the application process, the proposed method converted the data from discrete to continuous distribution.

8. Recommendations

1. Considering the proposed method for estimating the Cox PH model.

2. The use of other types of orthogonal wavelets, methods for estimating the threshold level, and the thresholding rules in estimating the Cox PH model.

3. Conducting future studies to estimate the parameters Weibull, Gomppertz, and Log-Logistic Regression model using Wavelet Shrinkage.

4. Using a Bayesian approach with Wavelet Shrinkage in estimation the Cox PH model.

Appendix

```
% Program
clc
clear all
%rng('default')
n=300;p1=3;p2=2;K=p1+p2;v1=.5;ru1=.8;ru2=-.8;
```

```matlab
for j=1:1000
x=rand(n,p1)*v1;  Censored=[ones(.6*n,1);zeros(0.4*n,1)];  e1=randn(n,1);
x1=zeros(n,1);    x1(1)=e1(1);    e2=randn(n,1);x2=zeros(n,1);x2(1)=e2(1);
beta1=[.5 .75 1]';beta2=[.5 -.5]'; h0=.1;
for i=2:n
   x1(i)=ru1*x1(i-1)+e1(i);  x2(i)=ru2*x2(i-1)+e2(i);
end
corrcoef(x2(1:n-1),x2(2:n)); plot(x2(1:n-1),x2(2:n),'.');
% The noise
lambda=0.5;mu=0;b=.25;u=rand(1,n);v=-log(u)/lambda;
z=randn(1,n);noise=(mu+b*sqrt(2*v).*z)'*10;
ht=h0.*exp((x*beta1+[x1 x2]*beta2))+noise;
[bCoxTD,logl,HCoxTD,stats] = ...
   coxphfit([x x1 x2],...
   ht,...
   'Censoring',Censored,...
   'Baseline',0);
AIC(j)=-2*logl+2*(K+1);BIC(j)=-2*logl+log(n)*K;
% proposed
XD                =                wdenoise(ht,'Wavelet','db13',
'DenoisingMethod','universal','ThresholdRule','soft');
[bCoxTD,logl,HCoxTD,stats] = ...
   coxphfit([x x1 x2],...
   XD,...
   'Censoring',Censored,...
   'Baseline',0);
AICw(j)=-2*logl+2*(K+1);BICw(j)=-2*logl+log(n)*K;
end
MAIC=mean(AIC); MAICw=mean(AICw);
MBIC=mean(BIC); MBICw=mean(BICw);
[MAIC MBIC MAICw MBICw]
```

References

1. Aako, O.L. and Are, S.O., (2020), comparison of Cox, Weibull, Gompertz and Log-Logistic regression models in survival analysis of hypertensive patients.

2. Chaubey, Y. P., Doosti, H., and Prakasa Rao, B. L. S. (2008), Wavelet based estimation of derivative of a density for negatively associated process. Journal of Statistical Theory and Practice, 2(3) 453-463.

3. Christophe Chesneau, Junke Kou and Fabien Navarro, (2019), Linear wavelet estimation in regression with additive and multiplicative noise, HAL Id: hal-01877543 https://hal.archives-ouvertes.fr/hal-01877543v2

4. Cox DR (1972), Regression models and life-table. (With discussion). J Royal Stati Soci B, 34, 187-220.

5. Donald B. Percival, Muyin Wang and James E. Overland, (2004), An Introduction to Wavelet Analysis with Application to Vegetation Time Series, University of Washington

6. Donoho D. L. and Johnstone I. M. , (1994) , Ideal Spatial Adaptation by Wavelet Shrinkage , Biometrika , 81 , 425-550

7. Eddy Armand Kwessi Nyandjou, (2011), Efficient Rank Regression with Wavelets Estimated Scores, dissertation submitted to the Graduate Faculty of Auburn University

8. Ekman, A., (2017), Variable selection for the Cox proportional hazards model: A simulation study comparing the stepwise, lasso and bootstrap approach.

9. Gail, M., Krickeberg, K., Samet, J., Tsiatis, A. and Wong, W., (2007), Statistics for biology and health.

10. Harrell, F.E., (2015), Cox proportional hazards regression model. In Regression modeling strategies (pp. 475-519). Springer, Cham.

11. Iolanda Lo Cascio , (2007), Wavelet Analysis and De-noising: new Tool for Economists, University of London.

12. Jeena Joy1, Salice Peter2, Neetha John3, (2013), de-noising using soft thresholding, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 3.

13. Jinru Wang, Zhenming Zhang, Xue Zhang, and Xiaochen Zeng (2020), Wavelet pointwise estimations under multiplicative censoring, International Journal of Wavelets, Multiresolution and Information Processing.

14. Katsuyuki Hagiwara, (2021), Bridging between soft and hard thresholding by scaling, Faculty of Education Mie University (Mie

Univ.)

15. Lawless JF, (1998), Parametric models in survival analysis. In encyclopedia of Biostatistics. Armitage P. Colton T. Wiley: New York, 3254-64.

16. Linyuan Li, Brenda MacGibbon, and Christopher Valenta, (2006), On wavelet estimation in censored regression, University of New Hampshire, USA.

17. Linyuan Li, Brenda MacGibbon, and Christopher Valenta, (2007), On the minimax optimality of wavelet estimators with censored data, Journal of Statistical Planning and Inference, University of New Hampshire, USA

18. Liu, Y., (2005), Incorporating time-dependent covariates in the Cox proportional hazards model: The LVAR approach (Doctoral dissertation, Purdue University).

19. Prakasa Rao, B. L. S. (1996), Nonparametric estimation of the derivatives of a density by the method of wavelets. Bull. Inform. Cyb., 28, 91-100.

20. Ramazan Gengay, Faruk Selguk and Brandon Whitcher, (2002), An Introduction to Wavelet and other Filtering Methods in Finance and Economics.

21. Rinku Saikia and Manash Pratim Barman, (2016), Comparing Cox Proportional Hazard Model and Parametric Counterpart in the Analysis of Esophagus Cancer Patient Data, IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 5, PP 16-21 www.iosrjournals.org

22. Rogério Porto, Pedro Morettin, Donald Percival, and Elisete Aubin, (2016), Wavelet shrinkage for regression models with random design and correlated errors, Braz. J. Probab. Stat. 30(4): 614-652, DOI: 10.1214/15-BJPS296

23. Royston, P. and Lambert, P.C., (2011), Flexible parametric survival analysis using Stata: beyond the Cox model (Vol. 347). College Station, TX: Stata Press.

24. Singh, R. and Mukhopadhyay, K., (2011), Survival analysis in clinical trials: Basics and must know areas. Perspectives in clinical research, 2(4), p.145.

25. Xing-cai Zhou, Ying-zhi Xu, and Jin-guan Lin, (2017), Wavelet estimation in varying coefficient models for censored dependent data, journal homepage: www.elsevier.com/locate/stapro.

26. Yogendra P. Chaubey, Hassan Doosti, Esmaeel Shirazi, B. L. S., and Prakasa Rao, (2010), Linear Wavelet-Based Estimation for Derivative of a Density under Random Censorship, Concordia University, Montr´eal, Canada, JIRSS, Vol. 9, No. 1, pp 41-51.