Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: DOI–10.1285/i20705948v7n1p26

A Simulation Study on Cox Regression with Weighted Estimations

By Ata Tutkun, Demirhan

April 26, 2014

Electronic Journal of Applied Statistical Analysis
Vol. 7, Issue 1, 2014, 26-36
DOI: DOI–10.1285/i20705948v7n1p26

A Simulation Study on Cox Regression with Weighted Estimations

Nihal Ata Tutkun* and Haydar Demirhan

Hacettepe University, Department of Statistics, Beytepe 06800 Ankara Turkey

April 26, 2014

Cox regression model has an important and glaring place in survival analysis. The key assumption is proportional hazards and violation of this assumption can invalidate outcomes of a study. Our approach will be to use Cox regression model with weighted estimation for a survival data set that includes both proportional and nonproportional hazards. We carried out a simulation study, considering different censoring rates, sample sizes, and tied observations. Simulation results are interpreted and discussed with the results obtained by traditional Cox regression model. Cox regression model with N-Prentice weighting function serves as a better model under all simulation scenarios.

keywords: Cox regression model, hazard ratio, log-rank test, nonproportional hazards, survival analysis, weighting function.

1 Introduction

Survival analysis has several application areas from medicine to marketing. Kaplan-Meier provides a method for estimating the survival curve, the log rank test provides a statistical comparison of two groups, and Cox regression model (CM) allows additional covariates to be included. Both of the latter two methods assume that the hazard ratio comparing two groups is constant over time (Bewick et al., 2004). The CM analyzes the survival data arising from studies where the response variable is the length of time taken to reach a certain end-point, often death. The CM is similar to a multiple regression model and enables analyzing the effects of covariates on the hazard function of time.

*Corresponding author: nihalata@hacettepe.edu.tr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

The CM is used extensively in clinical survival analysis. Its popularity is rooted in its simple interpretation, semi-parametric nature, mathematical elegance and its implementation in popular statistical software packages (Frankel and Longmate, 2002). It has a significant assumption of proportional hazards (PH) which means that the survival curves for two groups must have hazard functions that are proportional over time (i.e. constant relative hazard). Although the model is very useful when the assumption of proportionality holds, a special care is needed when it is violated (Perperoglou et al., 2007). If the PH assumption fails to hold for the treatment or for one or more covariates, the results of CM will be misleading. The hazard ratio for the particular covariate cannot be interpreted as a hazard ratio or relative risk; moreover, the regression coefficients of the other covariates may be biased as a result (Putter et al., 2005).

It is not uncommon to encounter nonproportional hazards in survival data applications, especially in long follow-up term studies. Several modeling techniques have been proposed for nonproportional hazards. Nonproportional hazards are generally modeled by extensions of Cox or parametric regression models. An alternative approach is Cox regression model with weighted estimation (WCM), which is suggested by Schemper (1992). The WCM has some advantages over models such as stratified CM and CM with time-dependent covariates, because there is no requirement of distributional assumption, additional parameters and determination of a strata variable.

In the study of Ata and Demirhan (2013), WCM is examined and weighting functions different from Schemper's suggestions are proposed (Schemper, 1992). In these studies, the weighting function is used only for the covariates with nonproportional hazards. The performance and effect of proposed and existing weighting functions are discussed via a simulation study. Their simulation study shows that WCM gives better results than the traditional CM for sample sizes 25 and 50 under 60% censoring rates. Under censoring rate of 60%, the WCM with weighting functions N-Altshuler, N-KM, N-Prentice, N-Prentice-Marek, N-Harris-Albert, N-Moreau, Tarone-Ware has better results than the traditional Cox regression model for $n = 25$; WCM with weighting functions N-Altshuler, N-KM, N-Prentice, N-Prentice-Marek, N-Moreau has better results than that of CM for $n = 50$.

In this study, we consider WCM for the survival data sets that include both proportional and nonproportional hazards. The aim of our study is to use different weighting functions not only for the covariates with nonproportional hazards but also for the covariates with proportional hazards. Additionally, we compare the behavior of examined models with CM.

Section 2 reviews Cox regression model with weighted estimation and investigate the proposed approach. The simulation study is given in Section 3. Results, discussions and some recommendations are given in Section 4.

2 Cox regression model with weighted estimations

Let T be the random variable representing time until the event of interest occurs. The distribution of T can be characterized by the hazard function which is defined as follows:

$$\lambda(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t | T \ge t)}{\delta t} \right\} \tag{1}$$

This is the probability that an individual alive at t experiences the event in the next period δt. If one or more measurements are available for each individual, we have a vector of covariates $\boldsymbol{x}$. The relationship between the hazard function of failure time and these covariates is analyzed by the following model proposed by Cox (1972):

$$\lambda(t; \boldsymbol{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}' \boldsymbol{x}) \tag{2}$$

where t denote failure time, $\boldsymbol{x}$ be a row vector of covariates of dimension p and $\boldsymbol{\beta}$ be a p-vector of covariate coefficients. $\lambda(t; \boldsymbol{x})$ is the hazard at time t for individual with covariate vector $\boldsymbol{x} = (x_1, x_2, \dots, x_p)$ and $\lambda_0(t)$ is an unknown function giving the hazard for the standard set of conditions, when $\boldsymbol{x} = \boldsymbol{0}$. This model is called Cox regression model (CM) and assumes that the hazard ratio between two subjects with fixed covariates is constant.

The simplest approximation to the likelihood function of CM and also WCM is given by Breslow (2004) in order to accommodate tied observations:

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{s}_j)}{\sum_{\ell \in R_{t_j}} \exp(\boldsymbol{\beta}' \boldsymbol{x}_\ell)} \tag{3}$$

Here, we suppose data are available for n individuals, amongst whom there are r district death times and $(n - r)$ right censored survival times. $t_1 < t_2 < \cdots < t_r$ are ordered death times. $R(t_{(j)})$, $j = 1, 2, \dots, r$, is the set of individuals who are alive and uncensored at a time just prior to $t_{(j)}$ and called as the risk set. $\boldsymbol{x}_\ell$ is vector of p covariates for the individual ℓ who is in the risk set. $\boldsymbol{s}_j$ is the vector of sums of each of the p covariates for those individuals who die at $t_{(j)}$. If there are d_j deaths at $t_{(j)}$, then the hth element of $\boldsymbol{s}_j$ is $s_{hj} = \sum_{k=1}^{d_j} x_{hjk}$, where x_{hjk} is the value of hth covariate, $h = 1, 2, \dots, p$, for the kth of d_j individuals who die at the jth death time (Collett, 2003).

The maximum likelihood estimates of the β-parameters in CM and also WCM can be found by maximizing log-likelihood function using numerical methods such as Newton-Raphson procedure. To obtain the estimates of β_j and information matrix $I(\hat{\beta})$, the first and second derivatives of log-likelihood function of WCM are given distinctly in Schemper (1992) and Ata and Demirhan (2013).

In previous studies, weighting function $(f_k(t_j))$ is taken as 1 for proportional hazards whereas different weighting functions are taken for nonproportional hazards. In this study, we use $f_k(t_j)$ different from 1 for the covariates with proportional hazards. We use weighting functions for all the covariates in the study, regardless of whether they have proportional hazards or not. Our first purpose is to compare WCM with CM and conclude which model is more appropriate under different scenarios. Our second aim is to check the performance of weighting functions with those in literature.

Some of the weighting functions suggested in Ata and Demirhan (2013) are considered in this paper and these are given in Table 1. In Table 1, weighting functions numbered by

(1)-(5) are the functions used for weighted log-rank tests in the literature, and weighting functions numbered by (1) (Gehan, 1965) and (2) are the functions used by Schemper (1992). Weighting functions numbered by (6)-(15) are the proposed functions formed from the estimation of survival functions and weighted log-rank tests in the literature.

Table 1: Weighting functions.

1	Gehan	n_j
2	N-KM	NS_j^{KM}
3	Moreau	$S_j^{PREN} = \prod_{i=1}^{j} \frac{n_i}{n_i+d_i}$
4	Tarone-Ware (TW)	$\sqrt{n_j}$
5	Andersen (Weighted Prentice-Marek)	$S_j^* n_j/(n_j+1)$
6	Altshuler	$S_j^{ALT} = \prod_{i=1}^{j} \exp(-d_i/n_i)$
7	Harris-Albert	$S_j^{**} = \prod_{i=1}^{j} \frac{n_i+d_i-1}{n_i+d_i}$
8	Peto-Peto (KM)	$S_j^{KM} = \prod_{i=1}^{j} \frac{n_i-d_i}{n_i}$
9	Prentice	$S_j^{*PREN} = \prod_{i=1}^{j} \frac{n_i}{n_i+1}$
10	Prentice-Marek	$S_j^{*} = \prod_{i=1}^{j} \frac{n_i-d_i+1}{n_i+1}$
11	N-Altshuler	NS_j^{ALT}
12	N-Harris-Albert	NS_j^{**}
13	N-Moreau	NS_j^{PREN}
14	N-Prentice	NS_j^{*PREN}
15	N-Prentice-Marek	NS_j^{*}

In Table 1, N denotes the number of individuals in the study, n_j denotes the individuals at risk and d_j denotes the individuals died within the interval $[t_j, t_{j+1})$. S_j^{ALT}, is the survival function estimate of Altshuler (1970), S_j^* is the survival function estimate of Prentice and Marek (1979), S_j^{**} is the survival function estimate of Harris and Albert (1991), S_j^{KM} is the survival function estimate of Kaplan and Meier (1958), S_j^{*PREN} is the survival function estimate of Prentice (1978), S_j^{PREN} is the survival function estimate of Moreu et al. (1992). These estimates are used in proposed weighting functions (Leton and Zuluaga, 2001).

In order to analyze the performance of traditional CM with WCM, we have carried out a simulation study that is presented in the next section.

3 Simulation Study

In this section, we carried out a simulation study to investigate CM and WCM with weighting functions used for all covariates. We would like to study the effect of the censoring rates and sample sizes on the estimates obtained from the models.

Two dichotomous covariates (X_1, X_2) are taken into consideration in this study. The variable X_1 is generated to have nonproportional hazards, whereas X_2 is generated to have proportional hazards. We use the simulation procedure of Schemper (1992). Survival times are generated from $Weibull(\alpha, \gamma)$ distribution and the setting of α and γ is chosen to get nonproportional hazards. To generate censoring times and determine the censored observations, we utilize from the studies of Persson (2005), and Jinnah (2007). Censoring times are drawn from uniform distribution $Uniform(0, k)$, where k is chosen to ensure desired censoring rates. We choose censoring rates (CR) that are $10\%, 30\%$, and 60% and values 25, 50, and 100 for sample size n. The simulated data is generated 1000 times.

A test based on the Schoenfeld partial residuals (Schoenfeld, 1982), which is a measure of the difference between the observed and expected value of the covariate at each time (Therneau and Grambsch, 2000), is used to control the PH assumption of the simulated data.

The weighting functions given in Table 1 are used for both proportional and nonproportional covariates. We obtain the bias and standard error (SE) of parameter estimates ($\hat{\beta}$ and $\hat{\beta}$) for CM and WCM. Then the sample mean square error (MSE) values are obtained to compare the performance of the models examined in the simulation study and the results are given in Table 2.

In the simulations, the Newton-Raphson algorithm is employed to obtain parameter estimates over the first and second derivatives of the likelihood function given in eq. (3) (Schemper, 1992). All computations for the simulation study are carried on the algorithm given in Appendix A over a script programmed on Matlab.

Table 2: MSE values for CM and WCM with different weighting functions.

		$n = 25$			$n = 50$			$n = 100$		
		$CR = 10\%$	$CR = 30\%$	$CR = 60\%$	$CR = 10\%$	$CR = 30\%$	$CR = 60\%$	$CR = 10\%$	$CR = 30\%$	$CR = 60\%$
CM		14.8584	16.8375	19.3013	12.8997	14.4886	17.0269	11.7432	12.6939	15.6196
WCM										
Weighting function	N-KM	14.9261	15.9492	17.6522	13.0763	14.1857	16.3596	12.0934	12.6653	15.3037
	Gehan	14.7885	15.8791	17.7489	13.1106	14.2708	16.3776	12.1232	12.716	15.3075
	Moreau	16.3089	17.3985	19.7985	13.6935	14.8434	17.1939	12.3549	12.9245	15.6725
	Tarone-Ware	14.4346	15.8489	17.74	12.8434	14.1663	16.3621	11.8542	12.6026	15.3066
	Andersen	16.5706	17.5916	20.0178	13.7608	14.8888	17.2375	12.3784	12.9373	15.6817
	Altshuler	16.3798	17.4227	19.813	13.7141	14.8508	17.1963	12.3634	12.9272	15.6729
	Harris-Albert	16.5154	17.8459	20.4139	13.7995	15.0537	17.4385	12.4301	13.0527	15.7667
	Peto-Peto (KM)	16.4615	17.4506	19.8305	13.7368	14.8592	17.1994	12.3724	12.9302	15.6736
	Prentice	15.7385	17.5082	20.1824	13.3188	14.831	17.3578	12.0261	12.881	15.7324
	Prentice-Marek	16.31	17.3999	19.7989	13.6945	14.8438	17.194	12.3557	12.9248	15.6725
	N-Altshuler	14.8796	15.9438	17.6506	13.0612	14.182	16.3595	12.0861	12.6634	15.3037
	N-Harris Albert	14.8866	15.9933	17.6422	13.0893	14.228	16.3756	12.1258	12.7213	15.3009
	N-Moreau	14.8403	15.9395	17.6492	13.0473	14.1786	16.3594	12.0792	12.6616	15.3038
	N-Prentice	14.4308	15.9049	17.6043	12.7458	14.1047	16.3596	11.7823	12.593	15.2932
	N-Prentice Marek	14.8406	15.9403	17.6492	13.0478	14.1787	16.3594	12.0797	12.6618	15.3038

As seen in Table 2, WCM for a part of weighting functions has better results than the CM in terms of MSE. The results in Table 2 are regarded as below:

For censoring rate 10%,

- When $n = 25$, WCM with weighting functions Tarone-Ware, N-Moreau, N-Prentice, N-Prentice-Marek, has better results than CM. Besides, N-Prentice has the minimum MSE within these weighting functions.
- When $n = 50$, WCM with weighting functions Tarone-Ware and N-Prentice has better results than that of CM. Besides, N-Prentice has the minimum MSE within the weighting functions.
- When $n = 100$, WCM with weighting function N-Prentice has better result than that of CM.

For all sample sizes, it is observed that CM has better results than WCM with weighting functions N-KM and Gehan. For censoring rate 30%,

- When $n = 25$, WCM with weighting functions N-KM, Gehan, Tarone-Ware, N-Altshuler, N-Harris Albert, N-Moreau, N-Prentice, N-Prentice-Marek has better results than CM. Besides, Tarone-Ware has the minimum MSE within these weighting functions.
- When n=50, same results with n=25 are obtained.
- When n=100, WCM with weighting functions N-KM, Tarone-Ware, N-Altshuler, N-Moreau, N-Prentice, N-Prentice-Marek has better results than CM. Besides, N-Prentice has the minimum MSE within these weighting functions.

For censoring rate 60%,

- When $n = 25$, $n = 50$, and $n = 100$, WCM with weighting functions N-KM, Gehan, Tarone-Ware, N-Altshuler, N-Harris Albert, N-Moreau, N-Prentice, N-Prentice-Marek has better results than CM. However, N-Prentice has the minimum MSE for $n = 25$ and $n = 100$ whereas N-Moreau has the minimum MSE for n=50 within the weighting functions.

Simulation results show that for all scenarios WCM with weighting function N-Prentice gives better results than CM and also WCM with Gehan and N-KM weighting functions. WCM with weighting function Tarone-Ware has better results than CM for sample size $n = 25$ and $n = 50$ under all censoring rates taken into consideration in this study.

4 A Real Data Application

A real data application is also carried out to see how these weighting functions work in reality. The breast cancer data set consists of 42 observations, of which 61.90% are

Table 3: MSE values for CM and WCM with different weighting functions.

	Models	$\hat{\beta}_1$	$\hat{\beta}_2$	$SE(\hat{\beta}_1)$	$SE(\hat{\beta}_2)$
CM		-0.2213	-0.5889	0.2387	0.2423
WCM					
Weighting function	N-KM	-0.2186	-0.6518	0.0021	0.0022
	Gehan	-0.2998	-0.6065	0.0023	0.0024
	Moreau	-0.2585	-0.5972	0.2629	0.2682
	Tarone-Ware	-0.2608	-0.5981	0.0236	0.0241
	Andersen	-0.2595	-0.5975	0.2656	0.271
	Altshuler	-0.2588	-0.5973	0.2631	0.2684
	Harris-Albert	-0.3013	-0.6079	0.2909	0.2985
	Peto-Peto (KM)	-0.2591	-0.5974	0.2633	0.2686
	Prentice	-0.2864	-0.603	0.2846	0.2914
	Prentice-Marek	-0.2588	-0.5973	0.2631	0.2683
	N-Altshuler	-0.2183	-0.6517	0.0021	0.0022
	N-Harris Albert	-0.2573	-0.6684	0.0024	0.0024
	N-Moreau	-0.218	-0.6515	0.0021	0.0022
	N-Prentice	-0.2432	-0.6619	0.0023	0.0024
	N-Prentice Marek	-0.2183	-0.6516	0.0021	0.0022

censored, is used in this section (Ata and Demirhan, 2013). The parameter estimates and standard errors are given by Table 3.

According to the standard errors of parameter estimates, WCM with N-KM, Gehan, Tarone-Ware, N-Altshuler, N-Harris Albert, N-Moreau, N-Prentice, N-Prentice Marek weighting functions have smaller standard errors than the CM. Hence, it is observed that results obtained for this data set are in rapport with the simulation results.

5 Conclusion

The Cox regression model is routinely used in clinical trials, but it may encounter serious difficulties with departures from the PH assumption, even when the departures are not readily detected by commonly used diagnostics (Frankel and Longmate, 2002). The assessment of PH assumption can be done by many numerical (a test based on inclusion of a time dependent covariate in the model (Cox, 1972), a test based on the Schoenfeld partial residuals (Schoenfeld, 1982), a test based on a comparison of different generalized rank estimators of the hazard ratio (Gill and Schumacher, 1987)) or graphical approaches (log-minus-log survival plots of survival functions, a plot of survival curves based on the Cox regression model and Kaplan-Meier estimates for each group, a plot of cumulative baseline hazards in different groups (Andersen et al., 1982), a plot of difference of the log cumulative baseline hazard versus time, a smoothed plot of the ratio of log-cumulative hazard rates time, a smoothed plot of scaled Schoenfeld residuals versus time and a plot of estimated cumulative hazard versus number of failures (Arjas, 1988)). None of these approaches are known to be superior in finding out nonproportionality. Interpreting graphical plots can be arbitrary. The conclusions are highly dependent on the subjectivity of the researcher. Thus, we can absolutely recommend the use of WCM with N-Prentice weighting function since it can be used for both proportional and non-proportional hazards. Further, WCM has no distributional assumption and has no need to examine PH assumption. Using WCM also eliminates the sorrows of which testing method of proportional assumption to use for the survival data.

Appendix A

The following algorithm is used for the computations in the simulation study.

Algorithm:

S1. Read the total number of simulations, sample size, censoring rate, parameters of Weibull distribution, real values of parameters, maximum iteration number, tolerance value and starting values for Newton Raphson algorithm.

S2. Repeat for the total number of simulations.

- **S2.1.** Randomly generate a data set.
- **S2.2.** Calculate all of the considered weights.
- **S2.3.** Repeat for each weight and no weight case.
  - **S2.3.1.** Find parameter estimates and corresponding variances by using Newton-Raphson algorithm.
  - **S2.3.2.** Calculate MSE value.
- **S2.4.** Store MSE value for each weight function.

S3. Calculate grand average of stored MSE values for each weight function over the total number of simulations.

S4. Write the MSE values for all weight functions.

Acknowledgements

The authors would like to thank the referees and the Editor for their useful and constructive comments which led to improve the manuscript.

References

Altshuler, B.(1970). Theory for the measurement of competing risks in animal experiments. *Mathematical Bioscience*, 6, 1-11.

Andersen, P.K., Borgan, G. and Keiding N. (1982). Linear nonparametric tests for comparison of counting process with application to censored survival data. *International Statistical Review*, 50, 219-258.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association*, 83, 204-212.

Ata, N. and Demirhan, H. (2013). Weighted estimation in Cox regression model: An application to breast cancer data. *Gazi University Journal of Science*, 26(1), 63-72.

Bewick, V., Cheek, L. and Ball, J. (2004). Statistical review 12: Survival analysis. *Critical Care*, 8, 389-394.

Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89-99.

Collett, D. (2003). *Modelling survival data in medical research.* Chapman and Hall.

Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34, 187-220.

Frankel, P. and Longmate, J. (2002). Parametric models for accelerated and long-term survival: a comment on proportional hazards. *Statistics in Medicine*, 21, 3279-3289.

Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52, 203-223.

Gill, R.D. and Schumacher, M. (1987). A simple test for the proportional hazards assumption. *Biometrika*, 74, 289-300.

Harris, E.K. and Albert, A. (1991). *Survivorship analysis for clinical studies.* Marcel Dekker.

Jinnah, A. (2007). *Inference for Cox's Regression Model via A New Version of Empirical Likelihood.* Unpublished Master of Science Thesis, Georgia State University.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.

Leton, E. and Zuluaga, P. (2001). Equivalence between score and weighted tests for survival curves. *Communications in Statistics-Theory and Methods*, 30(4), 591-608.

Moreau, T., Maccario, J., Lellouch, J. and Huber, C. (1992). Weighted log rank statistics for comparing two distributions. *Biometrika*, 79, 195-198.

Perperoglou, A., Keramopoullos, A. and Houwelingen, H.C. (2007). Approaches in modelling long-term survival: An application to breast cancer. *Statistics in Medicine*, 26(13), 2666-2685.

Persson, I. and Khamis, H. (2005). Bias of the Cox model hazard ratio. *Journal of Modern Applied Statistical Methods*, 4(1), 90-99.

Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika*, 65,167-179.

Prentice, R.L., Marek, P. (1979). A qualitative discrepancy between censored data rank tests. *Biometrics*, 35, 861-867.

Putter, H., Sasako, M., Hartgrink, H.H., van de Velde, C.J.H. and van Houwelingen, J.C. (2005). Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial. *Statistics in Medicine*, 24(18), 2807-2821.

Schemper, M. (1992). Cox analysis of survival data with nonproportional hazards functions. *The Statistician*, 41, 455-465.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards model. *Biometrika*, 69, 551-555.

Therneau, T.M., Grambsch, P.M. (2000). *Modelling survival data: Extending the Cox model*. Springer-Verlang.