Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2023;15(1):42-51

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2022-94594

Use of Survival Data in Multivariate Adaptive Regression Analysis: Simulation Study

Yaşam Süresi Verilerinin Çok Değişkenli Uyumlu Regresyon Analizinde Kullanımı: Simülasyon Çalışması

Merve TÜRKEGÜN ŞENGÜL[a], Gülhan TEMEL[a], İrem ERSÖZ KAYA[b]

[a]Department of Biostatistics and Medical Informatics, Mersin University Faculty of Medicine, Mersin, Türkiye
[b]Department of Computer Engineering, Tarsus University Faculty of Engineering, Mersin, Türkiye

This study was presented as an oral presentation at International Conference on Computational and Statistical Methods in Applied Sciences, November 9-11, 2017, Samsun, Türkiye.

ABSTRACT Objective: The multivariate adaptive regression splines (MARS) is very effective in order to model linear or non-linear relationships. The Cox regression residuals-based MARS model, which integrates Cox regression and MARS approaches, was created to assess the relationships between efficient risk factors on the survival. The purpose of this study is to introduce the Survival-MARS (SM) model which uses the Cox-Snell, Martingale, and deviance residuals. Also,our aim is to compare the performance of the models created with residuals at different sample sizes and correlation levels with the simulation study in order to determine the most effective residual type that can be used in the SM model. **Material and Methods:** Performances of SM models that use Cox-Snell,Martingale, and deviance residual types were compared at different sample sizes (n = 30, 100, 150, 250, 500, 1.000), with both no correlation (r = 0.00) and medium (r = 0.50) and high (r = 0.90) correlations between predictors. SM model performances were compared via minimum generalized cross-validation and the sum of mean squared error values.. **Results:** In all scenarios, SM models with Cox-Snell residuals have the best performance compared to other models established with other residuals. Martingale and deviance residuals were affected by high correlation and low sample sizes. **Conclusion:** In case of linear relationship between risk factors, SM models with Cox-Snell residuals are quite successful in explaining these relationship structures and enable the effects on the dependent variable to easily interpret.

Keywords: Multivariate adaptive regression splines; Cox regression; Cox-Snell residual; Martingale residual; deviance residual

ÖZET Amaç: Çok değişkenli uyumlu regresyon uzanımları tekniği [multivariate adaptive regression splines (MARS)], doğrusal ya da doğrusal olmayan ilişkileri modellemede oldukça etkilidir. Sağkalım üzerinde etkili risk faktörleri arasındaki bu tür ilişkileri değerlendirebilmek amacıyla her iki yöntemi birleştiren ve Cox regresyon artıklarına dayanan Survival-MARS (SM) modeli geliştirilmiştir. Bu çalışmanın amacı, Cox-Snell, Martingale ve sapma artıklarını kullanan Survival-MARS (SM) modelini tanıtmaktır. Ayrıca amacımız, SM modelinde kullanılabilecek en etkili artık türünü belirlemek için artıklarla oluşturulan modellerin, farklı örneklem genişliklerinde ve korelasyon düzeylerindeki performanslarını simülasyon çalışması ile karşılaştırmaktır. SM model performansları, minimum genelleştirilmiş çapraz geçerlilik ve minimum hata kareleri toplamı değerlerine bakılarak karşılaştırılmıştır. **Bulgular:** SM modellerinde ele alınan tüm senaryolar için Cox-Snell artıkları ile kurulan modellerin, diğer artıklarla kurulan modellere göre daha iyi performans gösterdiği görülmüştür. Martingale ve sapma artıklarının güçlü korelasyon ve küçük örneklem genişliklerinden etkilendiği gözlemlenmiştir. **Sonuç:** Risk faktörleri arasında lineer ilişki olması durumunda, Cox-Snell artıklı SM modeli bu ilişki yapılarını açıklamada oldukça başarılı olup bağımlı değişken üzerindeki etkilerinin de kolaylıkla yorumlanabilmesini sağlamaktadır.

Anahtar kelimeler: Çok değişkenli uyumlu regresyon; Cox regresyon; Cox-Snell artığı; Martingale artığı; sapma artığı

Correspondence: Merve TÜRKEGÜN ŞENGÜL
Department of Biostatistics and Medical Informatics, Mersin University Faculty of Medicine, Mersin, Türkiye
E-mail: merveturkegun@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 29 Nov 2022 ***Received in revised form:*** 23 Jan 2023 ***Accepted:*** 23 Jan 2023 ***Available online:*** 07 Mar 2023

Situations when more than one factor simultaneously affects a patient's prognosis are common in survival studies. Age, gender, tumor size, tumor stage, and kind of treatment (chemotherapy or radiotherapy) can all have an impact on survival time, particularly in cancer research. The development of an appropriate statistical prediction model is necessary for a thorough examination of these consequences. The expected event (death, relapse, recovery, etc.) may not be observed in some patients over the required follow-up period for a variety of reasons, making it impossible to determine their survival time in survival studies. These patients' data are referred to as censored data.[1] Classical approach are inadequate for censored data. Cox proportional hazard (PH) model is a popular method to use for development for modeling by taking into account covariates and censored observations that affect survival time.[2]

When continuous or categorical predictor variables are modeled at the same time in studies involving many variables, linear or non-linear relationships or interaction effects may occur between the variables. In this case, the Cox hazard function may be misleading, making accurate and effective estimations and interpretation difficult.[3] For this reason, a model was developed, combining Cox PH and multivariate adaptive regression splines (MARS) techniques to evaluation both survival time and complex relationships simultaneous called Survival-MARS (SM).[4,5]

The aim of our study is to introduce the SM model which uses the Cox-Snell (CS), Martingale, and deviance residuals. Also, performances of the SM model which have different types of residuals at different sample sizes and correlation levels are compared to find the best residual type for modeling SM.

MATERIAL AND METHODS

MARS

MARS is a non-parametric method developed by Friedman (1991) to investigate the interactions and non-linear relationships between the response variable and many predictor variables.[5] In the MARS, there is no special distribution assumption for the response and predictor variables, and all variable types can be used. MARS does not require the basic assumptions of linear regression methods. MARS is more flexible compared to linear regression methods because it is not affected by the multicollinearity and outliers.[6]

MARS explains non-linear relationships between independent variables with the help of piecewise linear regression. By separating the predictors into different regions, it creates regression slopes called fundamental functions (FF) for each region and provides approximate estimates. FFs are curves containing piecewise linear functions. The values that describe parts of piecewise linear regression are called “knots” for FFs and are denoted by *t*. MARS allows the slope of the regression line to change from one interval to another when two “knot” points intersect.[7] The general representation of the non-parametric regression model is expressed as, $y=f(x_1, x_2, \dots, x_p)+\varepsilon=f(x)+\varepsilon$ where y is response variable, p is explanatory variables, x is predictors ($x=(x_1, x_2, \dots, x_p)^T$), and ε is the error term. MARS obtains the f function by using FFs.

Piecewise linear functions are in the form of max (0, x-t) with a t-valued knot. It helps to create a flexible model using piecewise linear FFs, taking into account the situations in equation 1.

$$(x-t)_+ = \begin{cases} x-t, & x>t \\ 0, & others \end{cases} \quad \text{and} \quad (t-x)_+ = \begin{cases} t-x, & x<t \\ 0, & others \end{cases} \tag{1}$$

Finally, the MARS model is formed as an interaction of basic functions and interactions and is expressed by equation 2 where α_0 and α_m are FF coefficients and $B_m(x)$ is FF.

$$\hat{y} = \hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{2}$$

The MARS algorithm performs model prediction in two steps using forward selection and backward elimination methods. In the *forward selection method*; creates a model that contains the mean of the values of the response variable as a constant term. Then the FFs are determined and added to the model in pairs. This process is terminated when the maximum number of FFs is reached, which minimizes the error mean squares. In the *backward elimination method*, it works according to the effects of FFs on the model. The models with the maximum number of FFs created in the first step are trimmed. The FF is subtracted from the model until the FF with the minimum prediction error is reached. The model with the minimum mean squared error (MSE) is selected as the most appropriate model. Generalized cross-validation (GCV) statistics are used for the model selection criterion at this stage.[5] GCV is a statistic that helps reduce overfitting by penalizing a large number of FFs. GCV penalizes not only FFs but also knots. It allows the evaluation together both the residual error and model complexity by minimizing the MSE.[8] The model with the minimum GCV value is considered as optimal model.[9] GCV is calculated by (3),

$$\mathrm{GCV(M)} = \frac{1}{n}\frac{\sum_{i=1}^{n}[y_i - f(x_i)]^2}{\left[1-\frac{M+d*(M-1)/2}{n}\right]^2} \quad (3)$$

In equation 3, M is FF number, d is penalty parameter with a default value of 3, n is number of observations, $f(x_i)$ is predicted values of the MARS model.

COX PH MODEL

The Cox PH model is developed by Cox (1972) to determine the prognostic factors that affect survival. The Cox model investigates the association between covariates and the survival time of patients to predict hazard ratio. The hazard is defined as the occurrence of the failure event of interest. The Cox PH model is also a regression method that models the relationships between the hazard function and the covariates and is also called the PHs model. It is defined as the failure rate that continues over a short period of time where t is survival time $h(t)$ is hazard function. The Cox PH model for the regression coefficients ($\beta_1, \beta_2, \ldots, \beta_p$) of $h(t)$ that measures the relationship between covariates and survival, depending on p covariates ($x_1, x_2, \ldots, x_p$), is represented by equation 4. In the Cox PH model, h_0 *is* expressed as the initial hazard and it shows the hazard value when all x_i 's are equal to zero.[2]

$$h(t) = h_0(t) * \exp\{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\} \quad (4)$$

Covariates have a multiplicative effect on hazard in the Cox PH model. This situation reveals the PH assumption of the Cox regression model. According to the PH assumption; The hazard ratio does not change over time during follow-up. In other words, it means that the effect of all covariates on survival time is independent of time. In order to apply Cox regression, the PH assumption must be met.[10]

The Cox PH model is a semi-parametric method and it makes no assumptions and the survival time is censored. It differs from the linear regression model with this feature.[11]

RESIDUALS OF THE COX PH MODEL

CS Residuals

CS residuals is used as a measure of model fit in survival analysis. They are not distributed symmetrically around zero and do not take negative values. The CS (r_{Ci}) is expressed as (5),

$$r_{Ci} = \widehat{H}_i(t_i^*) = -\log \widehat{S}_i(t_i^*) \quad (5)$$

In equation 5, $\widehat{H_i}(t_i)$ and $\widehat{S_i}(t_i)$ is the cumulative hazard function's and survival function's predictive value of the individual i's on time t_i, respectively. When the model is appropriate, individual i is calculated at time t_i in a model-based estimation of an individual's survival function close to the true value of the individual's survival time $S_i(t_i)$, and r_{ci} shows an exponential distribution.[12]

Martingale Residuals

Martingale (M) residuals are used to determine the number of predictors that will enter the model and the model consistency by drawing the graphs against the predictors. Martingale residuals are the difference between the observed and expected number of failures of a subject. The values obtained by using the of CS residuals are multiplied by -1 and Martingale residuals (r_{Mi}) are obtained as (6),

$$r_{Mi} = \delta_i + r_{ci} \quad (6)$$

Martingale residuals are not symmetrical about zero. They take values in the range of -∞ and 1. They are unrelated to each other.[13,14]

Deviance Residuals

The graphs of the residuals are difficult to interpret because Martingale residuals have skewness makes due to their being non-symmetrically distributed around zero. For this reason, by applying a transformation to Martingale residuals, deviance residuals are obtained that are symmetrically distributed around zero and have approximately 1 standard deviation of the (- ∞,+ ∞) range. Deviance residuals are plotted against predictors. The deviance residuals, (r_{Di}) are calculated by (7),

$$r_{Di} = sgn(r_{Mi})[-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2} \quad (7)$$

In equation 7, the unit r_{Mi} *i-th* Martingale residual and $sgn(.)$ being the sign function of the Martingale residual.[13,15]

SM MODEL

SM model was developed to eliminate the negative effects of interaction structures, linear, and non-linear relationships between risk factors on survival.[3]

SM model is based on the use of residuals calculated by Cox regression analysis. Because these residuals contain both the censored data and the estimated survival time used in the residual calculations.[12] The residuals are defined as the model response variable in the MARS. Then, the MARS model is created with continuous or categorical predictors. Determining the risk factors affecting survival, linear and non-linear relationships are determined and interaction effects can be easily interpreted with the estimation model obtained.[2,3]

This study was planned as a simulation study. It is aimed to determine the type of residual that shows the best SM model performance using different residual types in both linear relations and with different sample sizes.

SIMULATION STEPS

Step 1: Generating the Correlated Variables

The binary variable X_1 was derived from the univariate distribution, and the continuous variables X_2 and X_3 were derived from the normal distribution with 0 mean and 1 standard deviation.

Sample sizes (n) are 30, 100, 150, 250, 500 and 1,000, and 6 different scenarios with 1,000 replications were created for each sample size.

Correlations between predictors were r=0.0 (no correlation), r=0.50 (medium), and r=0.90 (high).

Step 2: Data Generation for Survival Analysis

Survival time was derived from the Weibull distribution. Scale parameter was λ=10 and shape parameter was ν>0 for Weibull distribution. The censored data was generated randomly and censored from the right with a censoring rate of 25%. Survival status was derived from the Bernoulli distribution.

Step 3: Estimating of the Residuals

Cox PH regression analysis was performed on the generated data for each sample size. CS, Martingale, and deviance residuals were calculated and recorded for the SM model. All of these analyzes are performed using the in MATLAB 6.0 (The MathWorks, Inc.,United States) package program.

Step 4: The SM Model

The residuals obtained in step 3 are defined as the *"response variable"* in the MARS model, and for each sample size, the CS SM model ($SM_{C\text{-}S}$), Martingale SM model (SM_{Mar}), and Deviance SM model (SM_{Dev}) were created with STATISTICA version 13.3 (TIBCO Software Inc.,USA).

Step 5: Comparison of Performances

GCV and MSE values are calculated for the three SM models created in step 4. For GCV, the number of folds is taken as 10. Minimum GCV and minimum MSE values were used as model selection criteria.

RESULTS

The mean values of the MSE and GCV values obtained for each sample size from 1,000 simulations are presented in Table 1 when there is no correlation between the predictors. We observed that when the sample size increased, MSEs of all SM models were increased. The highest MSE values belong to the SM_{Dev} models, and the smallest MSE values belong to the $SM_{C\text{-}S}$ models. The MSE values of the $SM_{C\text{-}S}$ models were 0.30 and below in the uncorrelated condition.

According to Table 1, when the predictors had uncorrelated, GCV results were similar to MSE. While the GCV values of the SM_{Dev} and SM_{Mar} models increased, the GCV values of the $SM_{C\text{-}S}$ model decreased as the sample size increased. In the $SM_{C\text{-}S}$ model, GCV=0.325 for n=30 and GCV=0.308 for n=1,000. On the other hand, in the SM_{Dev} model, GCV=0.972 for n=30, GCV=1.135 for n=1,000, and in the SM_{Mar} model, GCV=0.434 for n=30 and GCV=0.496 for n=1,000. While there was no linear relationship between the variables, CS SM model was more successful than other SM models in both small and large sample sizes.

It is observed that MSE values had small increases with increasing sample size in the $SM_{C\text{-}S}$ for moderate correlation (r=0.50) between the predictors. MSE values for n=100-150-250 were equal or close to each other. On the other hand, GCV values decreased with increasing sample size. GCV was 0.329 for n=30, and GCV was 0.314 for n=1,000 (Table 2). For moderate correlation, the increase in sample size had a positive effect on the performances of the CS residual models.

TABLE 1: SM model performances in the uncorrelated condition.

r=0.00	Cox-Snell SM		Martingale SM		Deviances SM	
Sample size	**MSE**	**GCV**	**MSE**	**GCV**	**MSE**	**GCV**
30	0.290	0.325	0.402	0.434	0.898	0.972
100	0.286	0.309	0.459	0.475	1.048	1.084
150	0.289	0.311	0.470	0.483	1.070	1.103
250	0.292	0.310	0.479	0.489	1.095	1.119
500	0.297	0.310	0.488	0.495	1.114	1.130
1,000	0.301	0.308	0.493	0.496	1.126	1.135

SM: Survival-MARS; MARS: Multivariate adaptive regression splines; r: Correlation; MSE: Mean squared error; GCV: Generalized cross-validation.

TABLE 2: SM model performances for r=0.50.

r=0.50	Cox-Snell SM		Martingale SM		Deviances SM	
Sample size	**MSE**	**GCV**	**MSE**	**GCV**	**MSE**	**GCV**
30	0.298	0.329	0.406	0.439	0.904	0.976
100	0.302	0.318	0.460	0.476	1.041	1.080
150	0.302	0.316	0.470	0.483	1.074	1.101
250	0.306	0.316	0.480	0.490	1.098	1.115
500	0.308	0.315	0.488	0.494	1.116	1.127
1,000	0.311	0.314	0.493	0.496	1.128	1.134

SM: Survival-MARS; MARS: Multivariate adaptive regression splines; r: Correlation; MSE: Mean squared error; GCV: Generalized cross-validation.

TABLE 3: SM model performances for r=0.90.

r=0.90	Cox-Snell SM		Martingale SM		Deviances SM	
Sample size	**MSE**	**GCV**	**MSE**	**GCV**	**MSE**	**GCV**
30	0.298	0.329	0.406	0.439	0.905	0.979
100	0.300	0.316	0.459	0.475	1.045	1.081
150	0.302	0.316	0.470	0.484	1.072	1.100
250	0.304	0.315	0.479	0.489	1.096	1.116
500	0.308	0.315	0.487	0.493	1.116	1.127
1,000	0.312	0.316	0.494	0.497	1.128	1.134

SM: Survival-MARS; MARS: Multivariate adaptive regression splines; r: Correlation; MSE: Mean squared error; GCV: Generalized cross-validation.

In SM_{Mar}, both MSE and GCV values increased with sample size (MSE=0.406, GCV=0.439 for n=30 and MSE=0.493, GCV=0.496 for n=1,000). A similar situation was observed in SM_{Dev} (Table 2). It can be said that moderate correlation negatively affects the performance of models with Martingale and deviance residuals. Contrary to expectations, increasing the sample size did not improve the performance of the Martingale and deviance SM models.

When the three residual types were compared, the deviance SM models were most affected by the moderate correlation, while the CS SM models showed the best model performance. The weakest model performances belonged to the deviance SM models when compared to other models. Especially for small sample size, MSE=0.298 for $SM_{C\text{-}S}$ while it is 0.904 for SM_{Dev} (Table 2).

A high level of correlation caused an increase in MSE and GCV values (Table 3). When a high correlation (r=0.90) between the predictors, the best results were obtained from the $SM_{C\text{-}S}$ models. In the $SM_{C\text{-}S}$ model, the GCV value was calculated as very small even for a small sample size and high correlation. It was observed that $SM_{C\text{-}S}$ model was not affected by sample size and high correlation.

CS SM models performed best in all conditions. Minimum MSE (Figure 1A, Figure 2A) and minimum GCV (Figure 1B, Figure 2B) values were obtained from CS residual models for all correlated or uncorrelated cases.

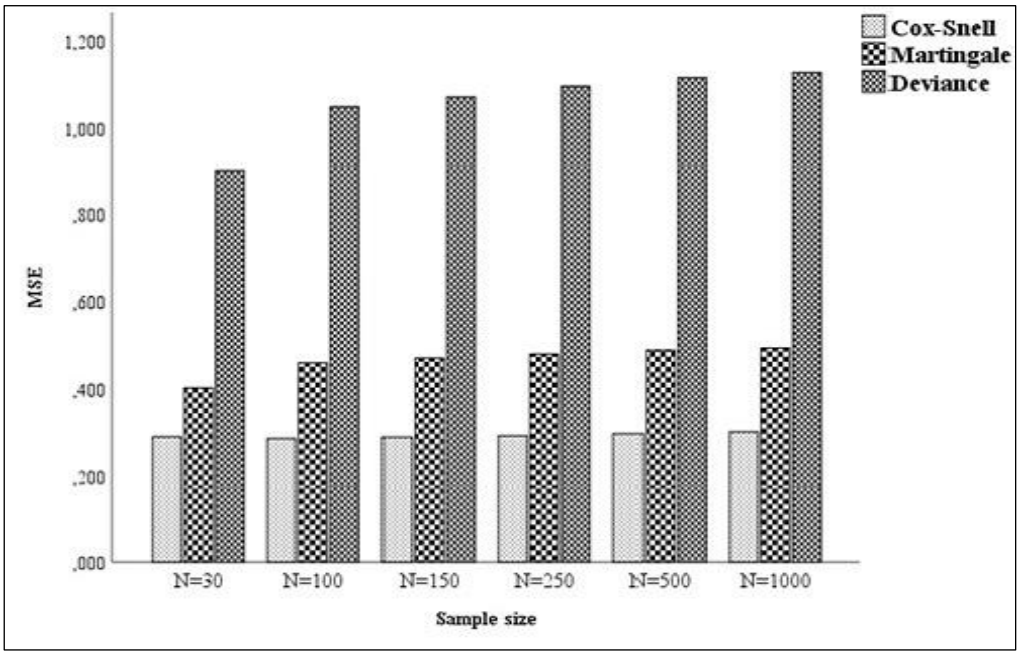

FIGURE 1A: MSE plot of Survival-MARS models for r=0.00.

MSE: Mean squared error; MARS: Multivariate adaptive regression splines.

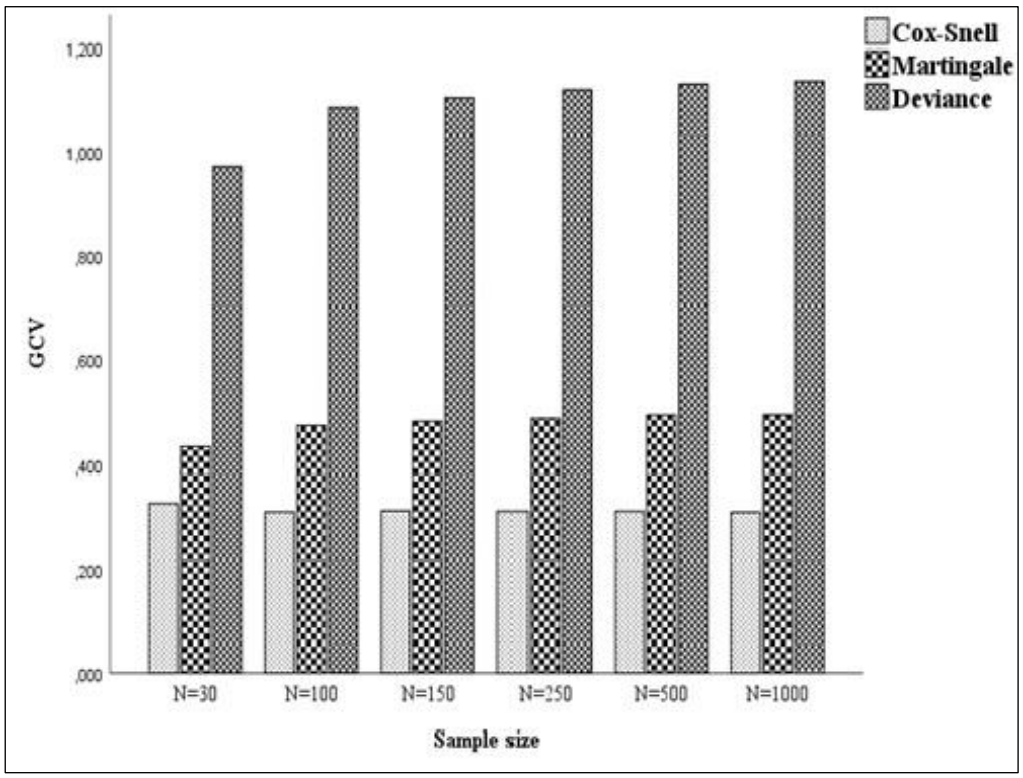

FIGURE 1B: GCV plot of Survival-MARS models for r=0.00.

GCV: Generalized cross-validation; MARS: Multivariate adaptive regression splines.

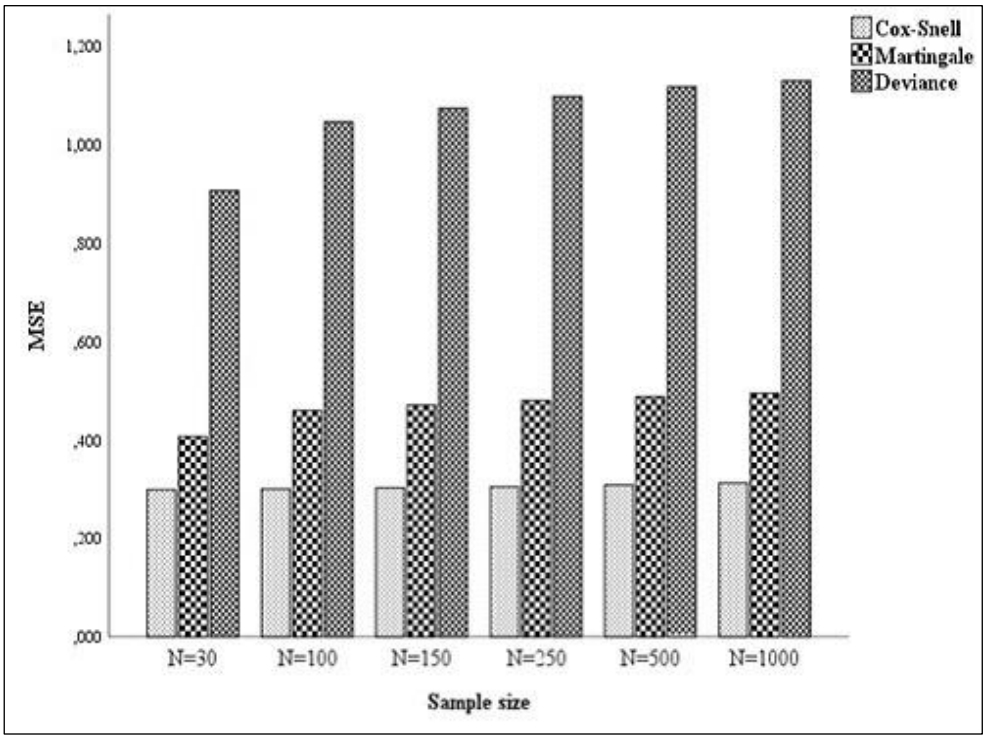

FIGURE 2A: MSE plot of Survival-MARS models for r=0.90.

MSE: Mean squared error; MARS: Multivariate adaptive regression splines.

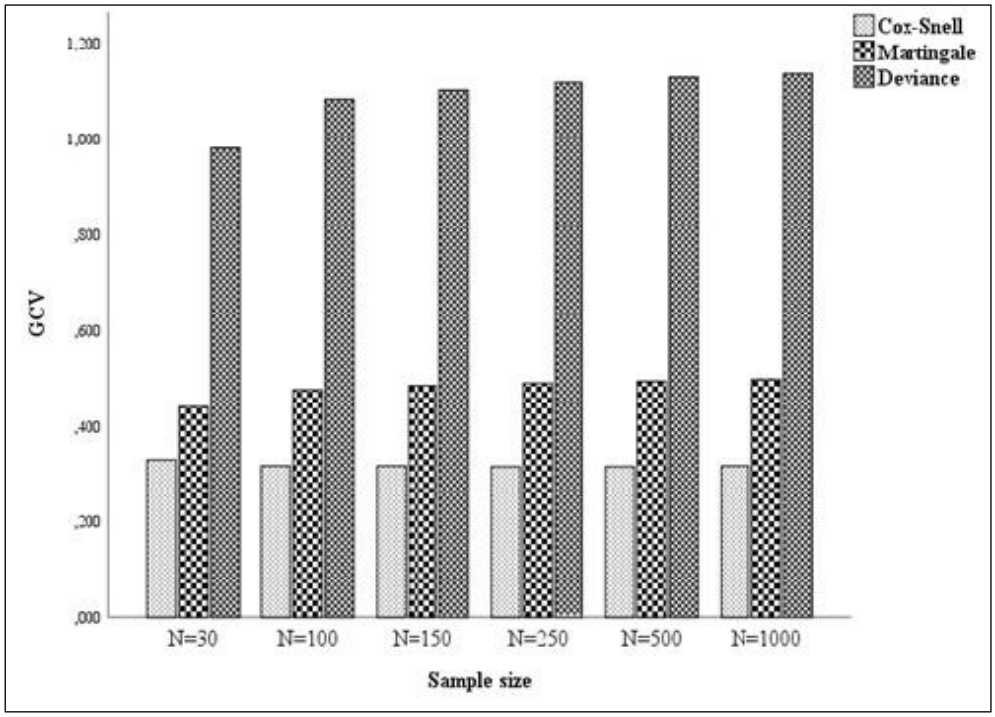

FIGURE 2B: GCV plot of Survival-MARS models for r=0.90.

GCV: Generalized cross-validation; MARS: Multivariate adaptive regression splines.

DISCUSSION

CS, Martingale, Deviance, Schoenfeld, and score residuals can be calculated by means of the Cox regression PH model. These residuals are plotted against predictors or time. They are used to examine the adequacy of the model, to decide on the variables to be included in the model and to be removed from the model, to determine whether the variable to be included in the model needs to be transformed and to check the PHs assumption.[11] In addition, residuals can also play a role in eliminating the deficiencies of statistical methods that are insufficient to explain complex structures among variables. In our study, the performances of SM models with CS, Martingale, and deviance residuals at different correlations and sample sizes were evaluated for cases with or without a linear relationship between the variables.

The lowest MSE and GCV values were obtained from CS SM model in the non-linear relationship. The highest MSE and GCV values belonged to the deviance SM models.

CS SM model had the lowest MSE and GCV values in a medium and high correlation. According to these results; it is seen that CS SM models are not affected by correlation and give successful results with increasing sample size.

The Martingale SM model showed the best model performance, secondly. The MSE values were calculated below 1 in all scenarios. The GCV values were also quite small and close to each other. It is thought that the Martingale SM model will produce valid and reliable results in cases CS residuals cannot be calculated.

Researchers may have a tendency to decrease variables when using multivariate statistical methods with small sample sizes to prevent multicollinearity. Generally, it is recommended that the required sample size should be 10 times the number of variables in order to get the most beneficial result from these methods.[16] According to the results from the SM models with three variables, it was observed that the sample size did not affect the MSE and GCV values, even if the correlation increased. For CS and Martingale SM models, while the correlation level was 0.90 in a small sample size (n=30); MSE was 0.298, 0.402; While correlation level was 0.00, MSE was 0.290 and 0.406, respectively. Similar results are also noticeable in GCV values. Since no similar study has been found with small sample sizes in the literature, we think that this study is a guide for future research.

Simulation studies have revealed the superiority of the SM models over the classical Cox PH model in non-linear relationships and interactions. In one study comparing the Martingale and deviance SM model using exponential and Weibull distributions, it was emphasized that the Martingale SM model is quite useful, superior, and powerful in explaining these effects.[3] The deviance SM model is the weakest to explain the relationships in this study like our study. In another study at different censoring rates (25%, 50%, 85%) and with exponential distribution, it was emphasized that the CS SM model was more effective than both the Martingale SM model and the classical Cox model. They also obtained similar results from the Weibull and Gompertz distributions.[17] Both the results obtained from our simulation study and similar studies showed that the distribution of survival time and censoring rates do not significantly affect the performance of the SM models. It draws attention that the CS SM models have the best model performance than the others.

The applicability of the SM model to real-life data is also available in various studies except from simulation studies. It observed that CS and Martingale residuals are often used in practice for SM models. These models were also compared with the Cox PH model. These studies gave similar results to simulation studies. According to this study, the MSE values of the CS SM model were smaller than the Martingale SM model and the Cox PH model.[18]

CONCLUSION

The SM model, which takes into account the complex relationships and interactions of predictors on survival times, facilitates the interpretation and intelligibility of the results. In addition, the use of this method is more flexible than Cox regression as it does not require robust assumptions. We recommend using the CS residual first and the Martingale residual second in the SM model rather than removing variables when there are numerous variables affecting the survival time.

The main framework in health research is to identify new or possible prognostic factors on a disease. If the researcher includes many variables that affect survival time and assesses each one's effects separately, assumes only linear relationships between predictors, and ignores interactions, the study results might not match predictions.[19] By eliminating the linear, non-linear, and interaction effects of predictors that have an impact on survival times, the recently created SM model presents a new perspective that makes interpretation easier.

It seems that the SM model was frequently used for non-linear and interactive effects in the literature. However, none of these studies evaluated the linear relationships for small sample sizes. We conducted a simulation study to find best SM model performance in small, moderate and large sample sizes. As a result, we observed that the sample size did not affect the SM model performances. This study is limited to the Weibull distribution with a 25% censor rate for survival time, three different correlations, and six different sample sizes. We recommend that these models be studied with different distributions and different censoring rates for linear relationships for as a future research subject.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Merve Türkegün Şengül, Gülhan Temel;* ***Design:*** *Merve Türkegün Şengül, Gülhan Temel;* ***Control/Supervision:*** Merve Türkegün Şengül, Gülhan Temel*;* ***Data Collection and/or Processing:*** *Merve Türkegün Şengül, Gülhan Temel, İrem Ersöz Kaya;* ***Analysis and/or Interpretation:*** *Merve Türkegün Şengül, Gülhan Temel;* ***Literature Review:*** *Merve Türkegün Şengül, Gülhan Temel;* ***Writing the Article:*** *Merve Türkegün Şengül, Gülhan Temel;* ***Critical Review:*** Merve Türkegün Şengül, Gülhan Temel.

REFERENCES

1. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. Br J Cancer. 2003;89(2):232-8. [Crossref] [PubMed] [PMC]
2. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B. 1972;34(2):187-220. [Crossref]
3. Monika Kriner. Survival analysis with multivariate adaptive regression splines [PhD thesis]. Munich: Ludwig Maximilian University of Munich; 2007. [Cited: Jun 9, 2022]. Available from: [Link]
4. LeBlanc M, Crowley J. Adaptive regression splines in the Cox model. Biometrics. 1999;55(1):204-13. [Crossref] [PubMed]
5. Friedman JH. Multivariate adaptive regression splines. The Annals of Statistics. 1991;19(1):1-67. [Crossref]
6. Temel Orekici G, Ankaralı H, Yazıcı AC. Regresyon Modellerine alternatif bir yaklaşım: MARS [An alternative approach to regression models: MARS]. Turkiye Klinikleri J Biostat. 2010;2(2):58-66. [Link]
7. Chou SM, Lee TS, Shao YE, Chen IF. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications. 2004;27(1):133-42. [Crossref]
8. Craven P, Wahba G. Smoothing noisy data with spline function, estimating the correct degree of smoothing by the method of generalized cross-validation. Numer Math. 1979;31:377-403. [Crossref]
9. Altınok G, Karagöz P, Batmaz İ. Learning to rank by using multivariate adaptive regression splines and conic multivariate adaptive regression splines. Computational Intelligence. 2021;37(1):371-408. [Crossref]
10. Kleinbaum D, Klein M. Evaluating the proportional hazards assumption. Survival Analysis: A Self-Learning Text. 3rd ed. New York: Springer; 1996. p.164-87. [Crossref]
11. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. Br J Cancer. 2003;89(3):431-6. [Crossref] [PubMed] [PMC]
12. Nardi A, Schemper M. New residuals for Cox regression and their application to outlier screening. Biometrics. 1999;55(2):523-9. [Crossref] [PubMed]
13. Collett D. Text in Statistical Science: Modelling Survival Data in Medical Research. 2nd ed. Boca Raton: Taylor & Francis Inc.; 2003.
14. Barlow WE, Prentice RL. Residuals for relative risk regression. Biometrika. 1988;75(1):65-74. [Crossref]
15. Therneau TM, Grambsch PM, Fleming TR. Martingale-based residuals for survival models. Biometrika. 1990;77(1):147-60. [Crossref]
16. Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 5. Baskı. Ankara: Detay Anatolia Akademik Yayıncılık LTD.ŞTİ.; 2017.
17. Dauda KA, Yahya WB, Banjoko AW. Survival analysis with multivariate adaptive regression splines using cox-snell residual. Anale Seria Informatică. 2015;3(2):25-41. [Link]
18. Irwansyah E, Stefani D, Bekti RD. Cox proportional hazard with multivarite adaptive regression splines to analyze the product sales time in E-commerce. Int J Appl Math Stat. 2015;53(5).109-15. [Crossref]
19. Austin PC, Fang J, Lee DS. Using fractional polynomials and restricted cubic splines to model non-proportional hazards or time-varying covariate effects in the Cox regression model. Stat Med. 2022;41(3):612-24. [Crossref] [PubMed] [PMC]