Bridge Deterioration Analysis Based on Censored Data: A Case Study in Yunnan Province

Xin He\textsuperscript{1*}, Wei Wang\textsuperscript{1,2}, Weidi Hu\textsuperscript{1} and Hanyu Zhang\textsuperscript{1}

\textsuperscript{1} Software School, Yunnan University, Kunming, Yunnan Province, 650500, China
\textsuperscript{2} Key Laboratory for Software Engineering of Yunnan Province, Kunming, Yunnan Province, 650500, China
* e-mail: 2205494777@qq.com

Abstract. Bridge deterioration reflects the change of bridge condition and it is caused by many factors. Predicting bridge deterioration as a research problem has attracted increasing attention around the world. Most methods contrived in previous work and the results are satisfactory. Despite their considerable success, these methods are weighed down by some problems. First, nature environment and climate factors are not considered. Second, they neglected the existence of censored data. Censored data means that the event outcome of interest can’t be observed exactly because of some reasons. It is usual in bridge deterioration prediction and will cause deviation between the prediction result and the actual situation. In this paper, we propose a method to tackle the censored data issue and add nature environment and climate factors into deterioration analysis. This method uses an indicator variable to identify censored data in parameter estimation. In order to ensure the objectiveness of experiment, data collected from the Yunnan Communications Investment & Construction Group LTD are used in this study. Empirical results show that different structure and material of bridge may have different effect on bridge condition. In addition, we confirmed that nature environment and climate can affect the deterioration rate of bridge comparing with other factors in Yunnan Province. The bridges in the areas with moderate temperature and large precipitation may deteriorate more slowly. These findings are important for bridge management and maintenance.

1. Introduction

Bridge deterioration analysis has attracted considerable attention for past ten years. Success in such analysis leads to the development of bridge maintenance research\cite{1, 2}.

In recent years, various machine learning algorithms adopted in deterioration analysis, such as Artificial Neural Network model\cite{3}, Markov chain model\cite{4}, Linear Regression model\cite{5} and fuzzy model\cite{6}, and produced many encouraging results. However, the advance of the deterioration analysis is slowed down by the two issues, without considering the nature environment factors and the nature of censoring of training data. These issues are described below.

(1) Natural environment and climate factors were not included in the study. The difficulty of bridge deterioration analysis is caused by two aspects: First, bridges are composed of several components, superstructure, substructure, decks, bearings, foundation, etc. Different components have different contribution on the bridge deterioration. Second, bridge deterioration is not only due to the age and structure, but also depends on complex environmental and climate factors. In order to get a more comprehensive analysis, all these factors should be considered in the study as far as possible.
(2) The nature of censoring of training data. Many models based on machine learning assume the data used to construct the deterioration analysis model is non truncated. However, the hypothesis is not always true. Censored data means that data was truncated because of some reasons. In bridge deterioration analysis, these reasons could be summarized as follows. First, some bridges were damaged by unnatural factors during the observation interval, such as artificial reconstruction, demolition, etc. Second, some bridges stop regular inspection during the observation interval. Third, the bridge did not deteriorate until the end of the observation interval. Fourth, the standard range of scores limits the assessment of bridges. The condition of bridge does not get a proper numerical value until the upper limit of the scoring criteria. The fourth case of censoring is usually neglected by previous studies. In statistics, censoring arises when an event of interest cannot be directly observed in the observation interval, which causes incomplete data\cite{7}. Most common problems encountered in the models are caused by the unobserved events, such as a large deviation between the prediction result and the actual situation because of data censoring\cite{8}. Censored data can’t be removed as exception values. Because it contains some important information, removing it will result in loss of sample size and information. If the information of these objects is lost, one-sided or unstable analysis results will be produced\cite{9}. Censored data also can’t be regarded as complete data. Otherwise, the bridge condition will be underestimated. If the bridge with very good condition in the censored data is lost, the prediction result will produce deviation.

In this paper, we add the data of nature environment and climate factors into the training data. Besides, aiming at the problem of censored data, Cox regression model is used to predict major factors that lead to the deterioration of bridge. Using Cox regression model, we can handle this incomplete data as censored data in the analysis instead of removing it. The partial regression coefficients show the magnitude and direction of the influence of the variable on the dependent variable. Cox regression also can filter variables and select variables that have more significant influence on the dependent variable by using stepwise regression.

The rest of paper is organized as follows. First, the current research states and progress of bridge deterioration analysis are reviewed. Second, necessary statistical background is briefly presented for censored data and Cox regression model. Then, the proposed method is presented. Afterward, the case study is described, followed by a presentation of empirical results, including the experimental setup and results analysis. Finally, the paper concludes with summaries and discussions.

2. Literature review

Numerous methodologies have been applied to bridge deterioration analysis. All these methods can be divided into two types: deterministic and stochastic\cite{10}.

The Markov chain is one of the stochastic methods which was used extensively in forecasting the performances of bridge infrastructures. As early as 1992, Cesare proposed methods\cite{11} for utilizing Markov chain in highway bridge deterioration research and bridge condition ratings were considered as Markovian states. Markov-chain model assumed that the future state is based on the present condition\cite{12}, each state can transfer to another at each time step according to the fixed transition probabilities. In bridge deterioration analysis, transition probability matrix was expressed as a matrix because that bridge condition is evaluated using several rating levels. The transition probability matrix is commonly calculated by statistical data of bridge conditions. Therefore, if the present bridge condition is known, the future condition at a certain time can be obtained. Up to recent years, Markov chain is still the mainstream model for bridge deterioration prediction. Li Li\cite{13} used a group of bridge condition data about ten years in Shanghai as training data, and proposed a Markov-chain model to predict the deterioration tendency of urban bridges in Shanghai. Although Markov chain is widely used in bridge deterioration analysis, most studies predict the transition of bridge condition, focus only on a particular bridge component or directly consider the bridge as a whole\cite{14} not consider the influence of nature environmental and climate factors to bridge, which may affect the accuracy of decision-making in bridge maintenance. In addition, previous work assumed that the data
used to construct the deterioration model without censoring. This hypothesis will lead to deviation in the experimental results.

Regression model is a deterministic method can quantify the impact of different factors on the final results. It is a predictive modelling technology, which is primarily used to find the best-fitted mathematical model, so that a dependent variable (objective) can be predicted from independent variables (predictors). It also serves a major purpose that describing the cause and effect relationship between dependent variable and independent variable. Generally, the potential determinants of bridge deterioration are identified first in bridge deterioration analysis, which were inputted as independent variables. And the condition rating was used as the dependent variable. The values of the coefficients estimated by regression analysis for the independent variables represent the influence extent to bridge deterioration. In reference[15], Yanev used regression model to analysis bridge deterioration under the only factors bridge age. Chase considered more factors including age, average daily traffic, precipitation, frequency of deicing, temperature range, freeze thaw cycles and type of bridge construction in reference[16], and they applied three different regression methods to capture the relationship between bridge condition and impact factors. They neglected the nature environment and climate factors. In addition, both of them also assumed that there is no censoring in the data used to construct the regression model.

Cox regression is a multivariate statistical method and has an ability to handle censored data[17]. Yang Ying Nan and Kumarswamy[18] used Cox regression analysis to predict the distribution of durations between changes in condition states. Censored data was mentioned and handled in their study, but they didn’t consider the case of censored data caused by score range limitation. Reference[19] described the data limitation in their study. There are two major problems. First, because of the subjective nature of bridge inspections, the same bridge was inspected by different inspectors could result in the same bridge got different ratings. Second, some newly constructed bridges are not given an initial condition rating of the best rating. They solved these problems by training inspectors, review of inspection reports, and spot checks of individual bridges. Similarly, reference[20] also found that bridge inspection observations are innately imprecise or fuzzy because of the subjective nature of bridge inspections and linguistic descriptions. A matrix-driven multiple fuzzy linear regression model was proposed to deal with a mixture of fuzzy data and crisp data. The two studies thought that these data problems were due to misapplication of rating criteria by bridge inspectors. However, when inspectors could assess bridge condition correctly according to rating criteria, we can find that the rating also can’t represent the actual bridge condition because of the score range limitation. This is one of the censoring cases, which is common in bridge deterioration analysis. However, the present studies neglected this case of censored data and the influence on bridge deterioration model.

Therefore, the aim of our study is to handle the censored data caused by score range limitation and an analysis method based on Cox regression was used in this paper.

3. Background

In this section, backgrounds are briefly presented for censored data and Cox regression model. More technical details are presented in [21] for more statistical treatments of this model.

3.1. Censored Data

Censoring represents an incomplete data structure, which means that the event outcome of interest can’t be observed exactly because of some reasons during the observation interval[7]. Censoring falls into three main categories, right censoring, left censoring and interval censoring[22]. Interval censoring means that the event outcome of interest is known only to lie within an interval [L, R] instead of being observed exactly. Left censoring means that the event outcome of interest occurred before the initial point of observation interval, but exact time is unknown. Right censoring is the most common type of censoring data in practical research. In right censoring, the initial point of observation interval is known. Generally, according to the difference of end point, right censoring can be divided
into three types: Type-I censoring, Type-II censoring and Type-III censoring. Type-I censoring means that the initial point is same. Except for those whose event outcome have already occurred, the end point of the rest of research objects is unified to a fixed time, and the event outcome of these objects cannot be observed, because they occur after the end point of the observation. For example, in bridge deterioration analysis, the minimum limit of bridge scoring standard is 0. Due to the limitation of the scoring criteria range, researchers cannot observe the outcome events of all bridges. The outcome event is that bridge score gets a proper numerical value. Higher numerical values cannot be given after reaching the maximum limit 100 of the scoring standard. The true values of bridge scores whose outcome event have not occurred are unknown, but not less than 100. The data type studied in this paper is Type-I censoring. When the observation does not end until enough event outcomes occur, the survival time of the individual without event outcome is unknown, which is called Type-II censoring. The type that different individuals have different initial point and censoring point is called Type-III censoring.

3.2. Cox Regression Model
Cox regression model was first proposed by Cox[23]. It is a semi-parametric regression model. Compared to parametric method, Cox regression model has an advantage that we don’t need to consider the distribution of data when we use it. Cox regression belongs to one of the analytic techniques used in survival analysis. At present, the application of survival analysis is no longer limited to the analysis of survival, but as an ideological method of dealing with censored data and has gradually formed a relatively complete new theoretical system[24]. The event outcome in survival analysis can be death or something we concern. Cox regression model has been extensively used in the medical field[25, 26], economics[27] and management field[28] since its introduction. It has shown the power of regression analysis for censored data[17] and achieved considerable success. Cox regression model takes event outcome and survival time (bridge score in this study) as dependent variables. An indicator variable was set to identify whether the event outcome occurs. It is easy to determine if the event outcome occurs, and it is customary to set the indicator variable as 1. If the event outcome does not occur, it will contain at least two situations. One is that the bridges have not had an outcome event until the end of the observation interval. The other one is that in the middle of the observation interval, the observation of some bridges were stopped for various reasons, resulting in the failure to observe the outcome. As long as the event outcome does not occur, the indicator variable of these bridge data was set as 0. For the incomplete data structure in bridge deterioration analysis, using cox regression model avoids the deviation caused by censored data and ensures the integrity of information.

4. Proposed Method
In this section, we introduce the process of bridge deterioration model construction based on cox regression model. It comprises three components: variable selection, model construction and parameter estimation. The whole process is shown in Fig. 1.
4.1. Variable Selection

Cox regression takes two variables as dependent variables. One is a classification variable. And another is a continuous variable. In this study, the event outcome and bridge score were taken as dependent variables.

An indicator variable c was set to represent the event outcome. If the event outcome occurs, then c=1, otherwise c=0. This indicator variable c also identifies whether the data is censored data. Then, the bridge condition score was defined as t. For a given individual i, we mark its score as ti.

In order to analyse the influence of impact factors on bridge condition score, all the impact factors were defined as independent variables. Assuming that the number of impact factors for score is k, all factors can be defined as x1, x2, ..., xk, respectively, and note as covariate vector form is X = (x1 x2 ... xk).

4.2. Model Construction

Cox regression is different from conventional regression analysis. It does not directly use bridge condition score as the dependent variable of regression equation. The influence of covariate variables on bridge condition score is reflected by the ratio of hazard function and baseline hazard function, in which the hazard function and baseline hazard function are unknown. The hazard function was defined as h(t), it measures the probability that the bridge gets a proper score. We can describe it as Eq.(1).

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq t < t + \Delta t | t \geq t)}{\Delta t}$$ (1)

At the same time, the baseline hazard function can be defined as \(h_0(t)\). It represents the form of a bridge's hazard function without effect of covariate variables. \(h_0(t)\) can be any function related to t. There is no any assumption for it in Cox regression model.

Supposing that the hazard function of bridges with various factors \(X = (x_1 x_2 ... x_k)\) represented by \(h_i(t, X)\). According to the proportional hazards and logarithmic linear hypothesis, the Cox regression model proposed for bridge deterioration analysis is defined as Eq.(2).

$$h_i(t, X) = h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$ (2)

where \(h_0(t)\) is the equal of \(h_i(t, X)\) when \(X = (0 0 .... 0)\); \(\exp(\beta_k x_{ik})\) denotes the effect of the X, the effect is the mortality rate from \(h_0(t)\) to \(h_0(t) \exp(\beta_k x_{ik})\), \(\exp(\beta_k x_{ik})\) can be regarded as a proportional constant. So Eq.(2) was called proportional hazards model. \(\beta_k\) are parameters (estimate value based on sample) describing the impact of various factors on hazards. When \(\beta_k = 0\), factor \(x_k\) has no effect on hazards; when \(\beta_k > 0\), factor \(x_k\) increases hazards; when \(\beta_k < 0\), factor \(x_k\) reduces hazards and becomes protective factor. Eq.(2) also can be converted into Eq.(3).

$$\ln \left[ \frac{h_i(t,X)}{h_0(t)} \right] = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$ (3)

The left part of the equation is the natural logarithm of the relative risk degree, and the right part of the equation is the linear function of the corresponding regression coefficients with covariate variables.

4.3. Parameter Estimation

Cox regression model does not make any assumptions about \(h_0(t)\). For two individuals, the hazard ratio of them is irrelevant to \(h_0(t)\). We can find that \(h_0(t)\) will cancel out in this expression from Eq.(4).

$$\frac{h_i(t, X)}{h_i(t, X)} = \frac{h_0(t) \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{h_0(t) \exp(\beta_1 x_{i'1} + \beta_2 x_{i'2} + \cdots + \beta_k x_{i'k})}$$

$$= \frac{\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{\exp(\beta_1 x_{i'1} + \beta_2 x_{i'2} + \cdots + \beta_k x_{i'k})}$$ (4)
Even in the case that $h_0(t)$ is unknown, we can still estimate parameters. The radio dependents only on regression coefficient $\beta$. This is the reason that Cox regression model is a semiparametric model. One of the advantages of Cox model is that it can obtain reliable estimation of regression coefficients and related statistical inferences without any assumptions about the functional form of the baseline hazard function $h_0$. We defined partial likelihood rather than traditional likelihood function when learning the model. A censoring indicator $\delta_j$ was used to identify the censored data, which equals 0 if the j-th bridge data is censored, otherwise is 1. This is the method of dealing with censored data. In addition, we set $R(t)$ to represent the hazard set of score $t$. Supposing that there are m events. Then we can define the corresponding conditional partial likelihood function as Eq.(5).

$$L_p(\beta) = \prod_{i=1}^{m} \left( \frac{e^{\beta x_j}}{\sum_{j \in R(t)} e^{\beta x_j}} \right)^{\delta_j} \tag{5}$$

We also can write the log-likelihood function as follows:

$$\log L_p(\beta) = \sum_{i=1}^{m} \delta_i [\beta x_j - \log(\sum_{j \in R(t)} e^{\beta x_j})] \tag{6}$$

The numerical methods such as iteration algorithm is usually used to derive the Eq.(6)[29], and obtain the maximum likelihood estimation value of the parameter $\beta$ and its standard error, then the likelihood ratio test, score test or Wald test are performed which determine whether each covariate has statistical significance. We can judge whether these factors are dangerous factors or protective factors for bridge deterioration according to the value of $\beta$. The statistical significance can help us to find the impact factors that have more significant influence on bridge deterioration from all factors.

5. Case Study

The implement of bridge deterioration analysis in Yunnan Province using Cox regression is described in this section. The data we use in experiment are processed and coded. The sample size is 2793, of which 746 data are right-censored. And the percentage of censored data is 27%. The experiment is implemented in R language.

5.1. Data Preparation

Multi-factors analysis on bridge deterioration need to consider more factors as much as possible. In this paper, we considered that the factors causing bridge deterioration are not only the age and structure of bridge, but also the environmental factors and climate. Yunnan Province is located in the southwestern frontiers of China. The climate of Yunnan Province is diverse, it crosses different climatic zones including temperate monsoon climate, subtropical monsoon climate and tropical rainforest climate. Most districts of Yunnan belong to the subtropical monsoon climate. The climate has remarkable features that small annual temperature difference, large daily temperature difference and distinct dry and wet seasons. The bridge data of Yunnan provide more comprehensive information for bridge deterioration analysis. Furthermore, we find that the bridge data collected from Yunnan have significant censoring, which is accord with our research purpose. Next, data preparation before experiment will be described.

5.1.1 Data Collection. Real world bridge condition data we use in this study were obtained through contacting directly the Yunnan Communications Investment & Construction Group LTD, China. The data include bridges on 14 highways and cover 16 districts. These districts are distributed in three climate zones, temperate monsoon climate, subtropical monsoon climate and tropical rainforest climate. In this study, bridges in subtropical monsoon climate were selected as representatives for Yunnan Province with 2793 data. The basic data is a sample of all relevant information for each bridge, each sample includes 12 attributes, there are underpass, deck pavement, deck width, bridge type, etc. In addition, considering the effect of temperature and precipitation, we add 6 attributes. There are annual average temperature, average minimum temperature, average maximum temperature, annual average precipitation, rainy days and snow days. Furthermore, an indicator variable $c$ was set to define the status of censoring. If the data is censoring, then $c=0$, otherwise $c=1$. 

IOP Conf. Series: Materials Science and Engineering 787 (2020) 012035 doi:10.1088/1757-899X/787/1/012035
5.1.2 Data Processing. First of all, there are some missing data in raw data. If we simply discard missing data, valuable information may be lost, and inferential power compromised[30]. In doing so, random forest (RF)[31] was applied to interpolate missing data. Then, we merge similar attribute values and remove unreasonable attribute values after discussion with experts in bridge engineering. In order to facilitate computer processing, we use real-number encoding. For example, in the attribute values of “Bridge Type”, “super bridge” is coded as 1, “large bridge” is coded as 2, “middle bridge” is coded 3, “small bridge” is coded 4.

5.1.3 Data normalization. Because different evaluation indicators have different dimension and dimension units, which will affect the result of data analysis. In order to eliminate the dimension effect between indicators, it is necessary to normalize the data to solve the problem of comparability between data indicators. So, we normalize the attributes whose attribute values are numerical values. The normalization method we used is linear normalization. Conversion function is shown in Eq.(7).

\[ x' = \frac{x - \min(x)}{\max(x) - \min(x)} \]  

(7)

After the processing, the overview of data pattern is shown in Table 1.

| Underpass | Deck Pavement | Deck Width | Bridge Type | ... | Annual Average Temperature | Snow Days | Score | c |
|-----------|---------------|------------|-------------|-----|-----------------------------|-----------|-------|---|
| 1         | 4             | 3          | 3           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| 1         | 4             | 3          | 3           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| 1         | 5             | 1          | 4           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| 1         | 5             | 1          | 4           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| 1         | 5             | 1          | 4           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| 2         | 5             | 1          | 3           | ... | 0.2105263157895             | 0.0612244897959 | 100   | 0 |
| ...       | ...           | ...        | ...         | ... | ...                         | ...        | ...   | ...|
| 3         | 3             | 1          | 4           | ... | 0.1973684210526             | 0.1020408163265 | 63.9  | 1 |
| 3         | 3             | 3          | 4           | ... | 0.1973684210526             | 0.1020408163265 | 60    | 1 |

5.2. Experimental Setup

5.2.1 Single Factors Analysis. In order to understand the determinant effect of each covariate variable on the model and which variable has impact on the bridge condition score, we do single factor analysis of 17 covariate variables by using Cox single factor analysis first. And we remove two covariate variables that fail to pass the significance test after the single factor analysis. The analysis results of covariate variable Deck Width are shown in Table 2 and Table 3.

| Variable name | Coefficient(β) | Standard Error(se) | z    | Pr(>|z|) |
|---------------|----------------|--------------------|------|---------|
| Deck Width-wide | -0.05168       | 0.11947            | -0.433 | 0.665 |
| Deck Width-narrow | -0.12599       | 0.09297            | -1.355 | 0.175 |

| Name             | Value  |
|------------------|--------|
| Likelihood ratio test | p= 0.4 |
| Wald test         | p= 0.4 |
| Score (logrank) test | p= 0.4 |

We do hypothesis test on the model and suppose the model is invalid, \( H_0: \beta = 0 \). According to the hypothesis test result of the model, with the significance level \( \alpha = 0.05 \), null hypothesis is invalid, i.e., the model only considering the impact factor Deck Width is not significant and doesn’t have statistical significance(when \( P<0.05 \), a model can be regarded as having statistical significance). Besides, according to the result of parameter estimation, variable Deck Width-wide does not pass the test of significance \( (P=0.665>0.05) \), so does variable Deck Width-narrow \( (P=0.175>0.05) \), which means that
this factor has little effect on bridge score. Therefore, the covariable variable Deck Width should be removed from the data set.

The analysis results of covariate variable Rainy Days are shown in Table 4 and Table 5.

Table 4. Parameter estimation of Rainy Days.

| Variable name | Coefficient(β) | Standard Error(se) | z  | Pr(|z|) |
|---------------|----------------|-------------------|----|--------|
| Rainy Days    | -0.10458       | 0.06544           | -1.598 | 0.11 |

Table 5. Hypothesis Test of Rainy Days.

| Name                  | Value          |
|-----------------------|----------------|
| Likelihood ratio test | p= 3e-12       |
| Wald test             | p= 9e-14       |
| Score (logrank) test  | p= 5e-14       |

Suppose the model only considering the impact factor Rainy Days is invalid, \( H_0 : \beta = 0 \). According to the hypothesis test result of the model, with the significance level \( \alpha = 0.05 \), null hypothesis is valid, i.e., the model is significant and has statistical significance (\( P<0.05 \)). But we find that the covariable variable Rainy Days does not pass the test of significance (\( P=0.11>0.05 \)) in parameter estimation. Therefore, the covariable variable Rainy Days should be removed from the data set.

In single factor analysis, 15 variables have significant influence on bridge score include Underpass, Deck Pavement, Bridge Type, Superstructure Form, Superstructure Material, Substructure Form, Substructure Material, Fundamental Form of Substructure, Support Form, Age, Annual Average Temperature, Annual Average Precipitation, Average Minimum Temperature, Average Maximum Temperature and Snow Days. All of these factors will be considered in multivariate Cox regression analysis.

5.2.2 Multi-factors Analysis. Cox single factor analysis has a problem that it ignores the possible interaction between variables. Synergetic effect could exist between some variables. When other factors are considered in the model, the influence of the factor on the score will be weakened or even become insignificant. So, multivariate Cox regression analysis is used in this paper. 15 variables are considered in Cox regression model.

The comprehensive evaluation index of the model and the result of goodness of fit test are shown in Table 6. The comprehensive evaluation index Concordance is 0.713. It shows that it has satisfactory indexes of discrimination. And the value of R-square that represents the goodness of fit is 0.315, which means that the expression equation has high reference value.

Table 6. The value of comprehensive evaluation index and goodness of fit.

| Name      | Value |
|-----------|-------|
| Concordance | 0.713 |
| R-square  | 0.315 |

Next, the hypothesis test of model is presented. Suppose the model is invalid, \( H_0 : \beta = 0 \). The result of hypothesis test is shown in Table 7.

Table 7. The result of hypothesis test.

| Name                  | Value          |
|-----------------------|----------------|
| Likelihood ratio test | p=<2e-16       |
| Wald test             | p=<2e-16       |
| Score (logrank) test  | p=<2e-16       |

According to the hypothesis test result of the model, with the significance level \( \alpha =0.05 \), null hypothesis is invalid, i.e., the model is significant and has statistical significance (\( P<0.05 \)). Then, the parameter estimation is presented in Table 8.
Table 8. Parameter Estimation of Cox Model.

| Parameter | Variable name                                      | Coefficient(β) | Exp(coefficient) | Standard Error(se) | z     | Pr(>|z|) | Signif. codes |
|-----------|---------------------------------------------------|----------------|------------------|--------------------|-------|----------|---------------|
| $X_1$     | Underpass-river                                   | 4.002e-01      | 1.492e+00        | 5.545e-02          | 7.218 | 5.29e-13 | ***           |
|           | Underpass-valley or gully                         | 2.681e-01      | 1.307e+00        | 8.788e-02          | 3.050 | 0.002286 | **            |
| $X_2$     | Deck Pavement-asphalt concrete                    | -1.323e+00     | 2.664e-01        | 3.298e-01          | -4.011| 6.05e-05 | ***           |
|           | Deck Pavement-cement concrete                     | -1.528e+00     | 2.170e-01        | 3.046e-01          | -5.016| 5.28e-07 | ***           |
| $X_3$     | Deck Pavement-asphalt concrete                    | -1.169e+00     | 3.107e-01        | 3.031e-01          | -3.857| 0.000115 | ***           |
| $X_4$     | Bridge Type-middle bridge                         | -8.775e-01     | 4.158e-01        | 4.369e-01          | -2.009| 0.044590 | *             |
| $X_5$     | Bridge Type-small bridge                          | -1.009e+00     | 3.644e-01        | 4.458e-01          | -2.264| 0.023570 | *             |
| $X_6$     | Superstructure Form-I type composite beam         | 1.319e+00      | 3.739e+00        | 3.710e-01          | 3.554 | 0.000379 | ***           |
| $X_7$     | Superstructure Form-hollow slab                   | 4.030e-01      | 1.496e+00        | 1.876e-01          | 2.149 | 0.031658 | *             |
| $X_8$     | Superstructure Form-box type                      | 5.674e-01      | 1.764e+00        | 2.636e-01          | 2.153 | 0.031334 | *             |
| $X_9$     | Superstructure Form-shape beam                    | 1.299e+00      | 3.667e+00        | 4.252e-01          | 3.056 | 0.002245 | **            |
| $X_{10}$  | Superstructure Form-prestressed hollow slab       | 6.893e-01      | 1.992e+00        | 1.616e-01          | 4.265 | 2.00e-05 | ***           |
| $X_{11}$  | Material-masonry coarse stone Superstructure      | 3.803e+00      | 4.486e+00        | 1.391e+00          | 2.734 | 0.006263 | **            |
| $X_{12}$  | Material-prestressed reinforced concrete Substructure | -4.511e-01   | 6.369e-01        | 1.667e-01          | -2.706| 0.006804 | **            |
| $X_{13}$  | Form-thin-walled abutment Substructure            | -6.794e-01     | 5.069e-01        | 3.196e-01          | -2.126| 0.033530 | *             |
| $X_{14}$  | Form-frame abutment (multi-column) Substructure   | -1.368e+00     | 2.547e-01        | 6.098e-01          | -2.243| 0.024901 | *             |
| $X_{15}$  | Form-ribbed column pier Substructure              | -1.539e+00     | 2.145e-01        | 2.512e-01          | -6.128| 8.88e-10 | ***           |
| $X_{16}$  | Form-buried abutment column pier Substructure     | -7.249e-01     | 4.844e-01        | 1.938e-01          | -3.739| 0.000185 | ***           |
| $X_{17}$  | Form-light abutment Substructure                  | -8.680e-01     | 4.198e-01        | 1.544e-01          | -5.621| 1.90e-08 | ***           |
| $X_{18}$  | Form-gravity Substructure                         | -4.276e-01     | 6.521e-01        | 1.419e-01          | -3.012| 0.002592 | **            |
| $X_{21}$ | abutment Substructure Form-gravity abutment column pier Substructure | -3.513e-01 | 7.038e-01 | 1.464e-01 | -2.400 | 0.016404 | * |
| $X_{22}$ | Form-pile column (T-shaped pier) Substructure | -1.017e+00 | 3.616e-01 | 2.587e-01 | -3.932 | 8.43e-05 | *** |
| $X_{23}$ | Form-pile column (column pier) Substructure | -8.979e-01 | 4.074e-01 | 1.548e-01 | -5.801 | 6.57e-09 | *** |
| $X_{24}$ | Material-reinforced mortar concrete Substructure Form of Substructure | -5.567e-01 | 5.731e-01 | 1.296e-01 | -4.296 | 1.74e-05 | *** |
| $X_{25}$ | expand foundation Substructure Form of Substructure | -1.784e+00 | 1.680e-01 | 7.184e-01 | -2.483 | 0.013034 | * |
| $X_{26}$ | & digging pile foundation Substructure Form of Substructure | -1.576e+00 | 2.067e-01 | 7.364e-01 | -2.141 | 0.032296 | * |
| $X_{27}$ | embedded foundation Substructure Form of Substructure | -1.519e+00 | 2.189e-01 | 7.701e-01 | -1.973 | 0.048523 | * |
| $X_{28}$ | gravel stone cushion Substructure Form of Substructure | -2.161e+00 | 1.152e-01 | 8.207e-01 | -2.633 | 0.008461 | ** |
| $X_{29}$ | Support Form-Φ250×300×57 Substructure Form | 1.260e+00 | 3.526e+00 | 4.552e-01 | 2.768 | 0.005639 | ** |
| $X_{30}$ | rectangular plate rubber bearing Support Form | -5.745e-01 | 5.630e-01 | 2.590e-01 | -2.218 | 0.026543 | * |
| $X_{31}$ | rubber bearing (plate, basin) Support Form | -6.035e-01 | 5.469e-01 | 2.945e-01 | -2.049 | 0.040443 | * |
| $X_{32}$ | round plate rubber bearing Support Form | -6.026e-01 | 5.474e-01 | 1.650e-01 | -3.653 | 0.000259 | *** |
| $X_{33}$ | Annual Average Temperature Average | 5.538e+00 | 2.542e+02 | 1.642e+00 | 3.374 | 0.000742 | *** |
| $X_{34}$ | Minimum Temperature | -1.538e+00 | 2.148e-01 | 3.245e-01 | -4.740 | 2.14e-06 | *** |
| $X_{35}$ | Annual Average Precipitation Snow Days | -4.533e+00 | 1.075e-02 | 8.522e-01 | -5.319 | 1.04e-07 | *** |
| $X_{36}$ | Snow Days | 1.979e+00 | 7.237e+00 | 6.082e-01 | 3.254 | 0.001138 | ** |
| $X_{37}$ | Age | 2.407e-01 | 1.272e+00 | 1.070e-01 | 2.250 | 0.024449 | * |
On the premise that the Cox regression model is valid, we can find that there are 37 variables have statistical significance (P<0.05) from the output of parameter estimation. On observing Table 8, it is clear that all the variables have stars in the last column. The star is significance code, where "***", "**", "*" represents P<0.001, P<0.01, P<0.05, respectively. The significance code indicates that a variable with more stars have more significant influence on the bridge score in the model.

In summary, empirical results show that the Cox regression model can generate in term of Table 8, and therefore, the model can be defined as Eq.(8).

\[
h(t, X) = h_0(t) \exp(4.002e - 01X_1 + 2.681e - 01X_2 + \cdots + 2.407e - 01X_{37}) \tag{B}\]

We can estimate whether the variable is a hazard factor or a protective factor according to Eq.(8). And the magnitude of the influence of the variable on the bridge score also can be presented. Because the influence of covariate variables on bridge score is reflected by the ratio of hazard function and baseline hazard function, the exp(coefficient) in Table 8 represents the change in the hazard ratio caused by a unit change of a variable. The higher the hazard, the higher the probability of low bridge score, and vice versa.

5.3. Result Analysis

On observing the empirical results, we can get some useful information for bridge management department. According to the influence on bridge score, these 17 factors selected by experts in bridge engineering are divided into three types, hazard factors, protective factors and factors without significant influence.

As we all know, the bridge condition will deteriorate rapidly with the passing of time. In other words, the longer the bridge is used, the lower the bridge score will be. This is also reflected in the experimental results. In addition to the age of bridge, other factors also have impact on bridge deterioration. First, we introduce the influence of bridge structure on bridge deterioration. In terms of Bridge Type, the deterioration rate of middle bridge is slower than large bridge and small bridge. Deck Width is a factor that doesn’t have significant influence on bridge score. Then, we find that Support Form is particularly special compared with other factors. When the attribute value of Support Form is Φ250×300×57, Support Form is a hazard factor which may cause faster deterioration of bridge. But when the Support Form is rubber bearing, it plays a protective role on bridge condition. In particular, using round plate rubber bearing as Support Form can significantly reduce the hazard of bridge deterioration. And the Superstructure Material is as special as Support Form. When the Superstructure Material is prestressed reinforced concrete, it is a protective factor. But when Superstructure Material is masonry coarse stone, it is a hazard factor. If Superstructure Form is shape beam or hollow slab, it will increase the hazard of deterioration. And using I type composite beam and prestressed hollow slab has a greater impact on deterioration hazard. In addition, we find some protective factors for bridge. When cement concrete and asphalt concrete are used on Deck Pavement, it has significant effect on reducing the hazard of deterioration. In order to reduce the hazard of bridge deterioration, these measures also can be adopted. Select reinforced mortar concrete as the Substructure Material. The Substructure Form can be ribbed column pier, buried abutment column pier, light abutment, pile column (T-shaped pier), pile column (column pier) or gravity abutment. And the Fundamental Form of Substructure can be gravel stone cushion, expand foundation, digging pile foundation and embedded foundation.

Beyond that, experiment proves that geographic location and climate also have impact on bridge deterioration. If the bridge is above river, valley or gully, it will increase the hazard of deterioration. Compare to valley or gully, the bridge over the river has higher hazard. In the aspect of climate, high temperature and snowfall will accelerate bridge deterioration. But it doesn’t mean that low temperature is good for bridge, the annual minimum temperature also cannot be too low. Furthermore, bridges in the places with large Annual Average precipitation will be in better condition. All in all, in the places where the temperature is moderate and the precipitation is large, the deterioration of bridge will be slower.
6. Discussion and Conclusion
In view of the two problems of the existing methods, we propose a method based on Cox regression model, which can not only comprehensively analyse the influencing factors, but also correctly handle the censored data. Using real world bridge data, the forecasting model is implemented and validated through investigations into the performance of Cox regression model. Empirical results show that the influence degree of factors on score. In addition, bridge management department can forecast the future bridge technical condition based on the assessment of these factors through the forecasting model and make a long-term maintenance planning.

Some remarks are provided as follows. First, other methods can be tried to fill the missing values in raw data to improve the accuracy of parameter estimation. Second, more feature selection methods can be explored. And more factors that may have influence on bridge condition can be added into the study, such as traffic volume. Moreover, in addition to Cox regression model, other approach might be investigated for handling censored data in real world bridge data. This is important for bridge deterioration analysis to get accurate prediction results.

Acknowledgments
The authors would like to thank the Yunnan Communications Investment & Construction Group LTD for providing the data used in this study. This work was supported by the National Natural Science Foundation of China under Grant No.61462092, and the Open Scientific Research Project of Yunnan Transportation Research Institute under Grant No.2013(B)01.

References
[1] Frangopol, D.M., Lin, K.Y., Estes, A.C.: (1997) Life-cycle cost design of deteriorating structures. Journal of structural engineering 123(10), 1390–1401.
[2] TSUDA, Y., KAITO, K., AOKI, K., KOBAYASHI, K.: (2006) Estimating markovian transition probabilities for bridge deterioration forecasting. Structural Engineering/Earthquake Engineering 23(2), 241s–256s.
[3] Huang, Y.H.: (2010) Artificial neural network model of bridge deterioration. Journal of Performance of Constructed Facilities 24(6), 195–201.
[4] Robelin, C.A., Madanat, S.M.: (2007) History-dependent bridge deck maintenance and replacement optimization with markov decision processes. Journal of Infrastructure Systems 13(3), 195–201.
[5] Hyman, W., Hughes, D., Dobson, T.: (1983) The least cost mix of bridge replacement and repair work on wisconsin’s state highways over time—a computer simulation. Technical Rep., WisDOT, Madison, Wis.
[6] Zhao, Z., Chen, C.: (2001) Concrete bridge deterioration diagnosis using fuzzy inference system. Advances in Engineering Software 32(4), 317–325.
[7] Rong, K.: (2008) Review of domestic research on censored data. In: Statistics & Information Forum, vol. 10.
[8] Nakat, Z.S., Madanat, S.M.: (2008) Stochastic duration modeling of pavement overlay crack initiation. Journal of Infrastructure Systems 14(3), 185–192.
[9] Miao, W., Yang: (2018) Cox regression analysis based on censored data. Statistics and Management 255(10), 26–29.
[10] Li, L., Li, F., Chen, Z., Sun, L.: (2016) Use of markov chain model based on actual repair status to predict bridge deterioration in shanghai, china. Transportation Research Record 2550(1), 106–114.
[11] Cesare, M.A., Santamaria, C., Turkstra, C., Vanmarcke, E.H.: (1992) Modeling bridge deterioration with markov chains. Journal of Transportation Engineering 118(6), 820–833.
[12] Ross, S.M., Kelly, J.J., Sullivan, R.J., Perry, W.J., Mercer, D., Davis, R.M., Washburn, T.D., Sager, E.V., Boyce, J.B., Bristow, V.L.: (1996) Stochastic processes, vol. 2. Wiley New York.
[13] Li, L., Sun, L., Ning, G.: (2014) Deterioration prediction of urban bridges on network level using markov-chain model. Mathematical Problems in Engineering 2014.

[14] Morcous, G.: (2006) Performance prediction of bridge deck systems using markov chains. Journal of performance of Constructed Facilities 20(2), 146–155.

[15] Yanev, B., Chen, X.: (1993) Life-cycle performance of new york city bridges. Transportation Research Record 1389, 17.

[16] Chase, S.B., Small, E.P., Nutakor, C.: (1999) An in-depth analysis of the national bridge inventory database utilizing data mining, gis and advanced statistical methods. Transportation Research Circular 498, 1–17.

[17] Therneau, T.M.: (1997) Extending the cox model. In: Proceedings of the First Seattle symposium in biostatistics. pp. 51–84. Springer.

[18] Yang, Y.N., Kumaraswamy, M.M., Pam, H.J., Xie, H.M.: (2013) Integrating semiparametric and parametric models in survival analysis of bridge element deterioration. Journal of Infrastructure Systems 19(2), 176–185.

[19] Veshosky, D., Beidleman, C.R., Buettow, G.W., Demir, M.: (1994) Comparative analysis of bridge superstructure deterioration. Journal of Structural Engineering 120(7), 2123–2136.

[20] Pan, N.F., Lin, T.C., Pan, N.H.: (2009) Estimating bridge performance based on a matrix-driven fuzzy linear regression model. Automation in construction 18(5), 578–586.

[21] Walters, S.J.: (1999) What is a Cox model? Citeseer.

[22] Cox, D.R.: (2018) Analysis of survival data. Chapman and Hall/CRC.

[23] David, C.R., et al.: (1972) Regression models and life tables (with discussion). Journal of the Royal Statistical Society 34(2), 187–220.

[24] Luo, S.: (2013) Method and application of survival analysis. Zhejiang Preventive Medicine 25(5), 29–31.

[25] D’Agostino, R.B., Lee, M.L., Belanger, A.J., Cupples, L.A., Anderson, K., Kannel, W.B.: (1990) Relation of pooled logistic regression to time dependent cox regression analysis: the framingham heart study. Statistics in medicine 9(12), 1501–1515.

[26] Sauerbrei, W., Schumacher, M.: (1992) A bootstrap resampling procedure for model building: application to the cox regression model. Statistics in medicine 11(16), 20932109.

[27] He, X., Chen, J.: (2014) An application of survival analysis in forecasting changes of stock index futures. Quantitative Economic, Technological and Economic Research pp. 116–126.

[28] Ting, L.: (2017) Research on Influencing Factors of Members’ Life Span in Virtual Community. Ph.D. thesis, Harbin University of Technology.

[29] Kalbfleisch, J.D., Prentice, R.L.: (2011) The statistical analysis of failure time data, vol. 360. John Wiley & Sons.

[30] Enders, C.K.: (2010) Applied missing data analysis. Guilford press.

[31] Breiman, L.: (2001) Random forests. Machine learning 45(1), 5–32.