

Research Article

Prediction of Bridge Component Ratings Using Ordinal Logistic Regression Model

Pan Lu ¹, Hao Wang ², and Denver Tolliver¹

¹Upper Great Plain Transportation Institute, North Dakota State University, Fargo, ND 58108, USA

²Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

Correspondence should be addressed to Hao Wang; hw261@soe.rutgers.edu

Received 2 January 2019; Revised 3 March 2019; Accepted 8 April 2019; Published 16 April 2019

Academic Editor: Jian Li

Copyright © 2019 Pan Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prediction of bridge component condition is fundamental for well-informed decisions regarding the maintenance, repair, and rehabilitation (MRR) of highway bridges. The National Bridge Inventory (NBI) condition rating is a major source of bridge condition data in the United States. In this study, a type of generalized linear model (GLM), the ordinal logistic statistical model, is presented and compared with the traditional regression model. The proposed model is evaluated in terms of reliability (the ability of a model to accurately predict bridge component ratings or the agreement between predictions and actual observations) and model fitness. Five criteria were used for evaluation and comparison: prediction error, bias, accuracy, out-of-range forecasts, Akaike's Information Criteria (AIC), and log likelihood (LL). In this study, an external validation procedure was developed to quantitatively compare the forecasting power of the models for highway bridge component deterioration. The GLM method described in this study allows modeling ordinal and categorical dependent variable and shows slightly but significantly better model fitness and prediction performance than traditional regression model.

1. Introduction

The highway bridge system is generally considered an essential part of the US transportation infrastructure. The efficient use of public funds for repairing and maintaining bridges requires an effective bridge asset management framework. Transportation management agencies worldwide have begun to adopt bridge management systems (BMS) to determine the optimum future bridge maintenance, repair, and rehabilitation (MRR) strategy at the lowest possible life-cycle cost based on forecasted bridge conditions [1–4]. Use of various forecasting models has played a critical role in predicting future bridge conditions for decision makers.

In the United States, highway bridge ratings typically consist of three major components: deck, superstructure, and substructure. The current method for monitoring them relies heavily on visual inspections which only take into account the observed physical health of the bridge. During visual inspections, a condition rating of the three major components is given on an integer scale of 0 to 9 with 8 equal interim levels. On this scale, 0 is failure and 9 is near-perfect condition

[5]. Bridge components deteriorate as a result of operating conditions and external environmental loads [6]. Because of the importance of these components for normal operation and safety, prediction models for component conditions are routinely developed to assess the condition of bridges for a given future time span.

A review of the literature reveals that many researchers have used various models to predict the future condition rating of bridges ([2–4, 7–15]; Lu and Zheng 2017). These prediction models include straight-line extrapolation, linear regression, Markovian, nonlinear regression, logistic regression models, artificial neural networks, Bayesian network, simulation, and some data mining based algorithms [7].

Applications of deterioration models such as Markov-chains and simulation are gaining popularity in forecasting bridge condition ratings; however they are limited by their inability to provide specific information on the deterioration of an individual infrastructure element [16]. In addition, the assumption in the Markovian method that probabilistic deterioration in a given period is independent of history can be unrealistic for bridges [9]. Artificial intelligence such

as neural networks is advantageous because of their ability to model nonlinearities automatically. Neural networks can handle binary categorical inputs by using 0/1 inputs, but it would be difficult to handle multiple categories that are ordinal in nature. Moreover, neural networks are more of a “black box” method that produces results that are difficult to interpret [16]. Multiple linear regression’s simplicity and explanatory relationship explains its popularity in literature, but this method may not be appropriate to model bridge condition because it does not take into consideration the ordinal discrete dependent variable. Use of multiple linear regression in that case will result in a violation of the normality assumption [9, 16]. Logistical methods such as the probit model allow the capture of the latent nature of infrastructure performance and incremental discrete dependent variables but do not adequately account for the multilevel discrete and ordinal nature of bridge ratings [9].

Multinomial regression is a variant of nonlinear regression that is capable of handling discrete dependent variables with multiple levels. However, bridge condition ratings are commonly represented as variables that are both discrete and ordinal in nature. In multinomial logistic regression, values of the dependent variable do not indicate any order or ranking. Ordinal logistic regression is an extension of multinomial regression that is believed to be theoretically appropriate and practically feasible for modeling bridge component rating changes. Those logistic models have been widely adopted in modeling discrete choices in motor vehicle crash severity and, to a lesser degree, in pipeline deterioration and wastewater utility deterioration [16]. However, use of the method to model bridge component or element rating changes has rarely been found in previous studies [17]. Madanat, Mishalani, and Ibrahim [17] presented an ordered probit method for the estimation of infrastructure deterioration models and associated transition probabilities from condition variables.

Moreover, the accuracy of the decision-making relies heavily on the outcomes of a reliable bridge condition forecasting model ([2, 15]; Lu and Zheng 2017). Many recent researches focus on improving model forecasting accuracy and shed some light on improving forecasting accuracy [2, 4, 12–15]. However, many of the researches ignore the forecasting reliability and provide incomplete picture regarding forecasting accuracy. Accuracy reported through previous research are often the statistical relative closeness measurement of model estimation to the actual condition [12, 18]; those measurements are critically important to demonstrate statistical soundness of the model’s forecasting power; however, they do not provide the full picture of the forecasting capability. For example, for discrete values such as bridge component ratings, estimation closeness along with exact estimation and estimation within certain rating difference will provide more complete pictures. Moreover, many previous researches ignore the forecasting reliability issue [18]. Their forecasting model can perform really well with certain data or dependent variables but work really poor with others. Thomas and Sobanjo [12] proposed a semi-Markov chain deterioration model, working really well with pourable joint seal element condition forecasting with relative closeness of 0.981 as 1 being perfect estimation,

but working really poor with reinforced concrete abutment element condition forecasting with relative closeness of 0.154.

In this research, the authors will evaluate the model forecasting capability based on various measurements including relative closeness measurements and exact accurate. Moreover, this research will validate the model forecasting power with not only in-sample data but also external data validation for three bridge component ratings.

2. Objective

In this study, an ordinal logistic regression method was developed to predict network-level bridge component ratings with North Dakota 2012 NBI data. A multiple linear regression model was also developed with the same data set as a reference for comparing model fitness and forecasting skill. The model is not perfectly suited for handling ordinal data as stated earlier; however it can be used for comparison since this type of model is popular within engineers and are straight forward to develop and use. Five criteria were used to evaluate and compare the two models: prediction error, bias, accuracy, out-of-range forecasts, Akaike’s Information Criteria (AIC), and log likelihood (LL). The developed model was validated with North Dakota 2013 and 2014 NBT data. The application of the model for predicting MAP-21 bridge performance indicator was conducted and discussed.

3. Ordinal Logistic Regression

Ordinal logistic regression is used to model the relationship between an ordered multilevel dependent variable and independent variables. In the modeling, values of the dependent variable have a natural order or ranking. One example of ordinal variables is bridge component ratings (ranging from 0 to 9, with 0 being fail and 9 being near-perfect). When the response categories are ordered, in ordinal logistic regression model, the event being modeled not only is having an outcome in a particular category but also preserves information about response categories which are ordered. Ordinal logistic regression models, also known as proportional odds models, utilizing proportional odds, have the following general form [19] shown in

$$\ln \left(\frac{\gamma_i^{(j)}}{1 - \gamma_i^{(j)}} \right) = \ln \left(\frac{P(Y_i \leq j | x_1, x_2, \dots, x_p)}{1 - P(Y_i \leq j | x_1, x_2, \dots, x_p)} \right) \quad (1)$$

$$= \tau_j - (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

where Y is response variable with k ordered categories;

$j=1,2,\dots,k-1$;

$\gamma_i^{(j)}$ is cumulative probability $P(Y_i \leq j) = P(Y_i = 1) + P(Y_i = 2) + \dots + P(Y_i = j)$ for $j=1,2,\dots,k-1$. Note $\gamma_i^{(k)} = P(Y \leq k) = 1$, so it should not be modeled;

Y_i are dependent observations which are statistically independent $i=1,2,\dots,n$;

x_1, x_2, \dots, x_p are p explanatory variables;

TABLE 1: Condition ratings used in the National Bridge Inventory (NBI).

Code	Meaning	Description
9	Excellent	As new
8	Very Good	No problems noted.
7	Good	Some minor problems.
6	Satisfactory	Structural elements show some minor deterioration.
5	Fair	All primary structural elements are sound but may have minor section loss, cracking, spalling or scour.
4	Poor	Advanced section loss, deterioration, spalling or scour.
3	Serious	Loss of section, deterioration, spalling or scour has seriously affected primary structural components. Local failures are possible. Fatigue cracks in steel or shear cracks in concrete may be present.
2	Critical	Advanced deterioration of primary structural elements. Fatigue cracks in steel or shear cracks in concrete may be present or scour may have removed substructure support. Unless closely monitored it may be necessary to close the bridge until corrective action is taken.
1	Imminent Failure	Major deterioration or section loss present in critical structural components or obvious vertical or horizontal movement affecting structure stability. Bridge is closed to traffic but with corrective action may put back in light service.
0	Failed	Out of service, beyond corrective action.

Source: United States Department of Transportation. Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges. Washington, D.C., 1995, page 38.

$\beta_1, \beta_2, \dots, \beta_p$ correspond to the regression coefficients for the respective independent variables;

τ_j are the cut-off points between categories.

Multinomial logit models do not consider proportional odds and ignore ordered response categories. For k possible outcomes, running $k-1$ independent binary logistic regression models in which one outcome, say k , are chosen as a reference and then the other $k-1$ outcomes are separately regressed against the reference outcome. The general form is followed by the following equation:

$$\ln\left(\frac{P(Y_i = j)}{P(Y_i = k)}\right) = \beta_j x_i \quad \text{where } j = 1, 2, \dots, k-1 \quad (2)$$

The restriction of ordinal regression originates from the proportional odds assumption even though ordinal regression takes care of ordinal relationship between levels of the dependent variable [16]. The proportional odds assumption is that β is independent of j . In other words, the effects of independent variables, β_s , are constant between different levels of the dependent variable. The proportional odds assumption can be tested by using a likelihood ratio score test to determine whether allowing the effects of independent variables to change will result in significant improvements in model fitness [16]. If the proportional odds assumption is not met, there are still several options, such as using the partial proportional odds model [20]. Our models meet the proportional odds assumption, possibly because of the large sample size and continuous latent response. The proportional odds cumulative-logit model acts well with its connection to the idea of a continuous latent response. Bridge condition is actually a categorized version of a latent continuous variable.

The 9-point scale is a coarsened version of a continuous variable indicating degree of component condition. The continuous scale is divided into 9 regions by 9 cut-points: 0-9. If we have normal errors rather than logistic errors, or in other words when an error term is a random error from a logistic distribution with mean zero and constant variance, the coarsened version of a continuous variable will be related to the independent variables by a proportional odds cumulative-logit model. It worth mentioning that the 9-point scale of bridge component ratings is subjective and there is a great need to model the relationship between the inspection rating and the actual condition of the bridge components. However, in this research the main focus is to demonstrate the forecasting improvement of the proposed model with 9-point scale measuring components due to the data availability.

4. National Bridge Inventory Database

The National Bridge Inventory (NBI) ASCII database is a unified database compiled by the Federal Highway Administration (FHWA) for all bridges and tunnels in the United States that have public roads passing above or below them [21]. The database provides the most comprehensive bridge information in the United States. Detailed information regarding NBI data can be found in the FHWA NBI reference report [22]. The data in the NBI is collected by state highway agencies and reported to FHWA annually/biennially.

As stipulated in the National Bridge Inspection Standards, bridges are inspected at least once every 24 months. During these inspections, the conditions of the three major bridge components (deck, superstructure, and substructure) are rated using a standard scale developed by Federal Highway Administration (Table 1). One can tell easily that

bridge component ratings are ordinal discrete data from Table 1. North Dakota 2012 NBI data is selected for model formulation and North Dakota 2013 and 2014 NBI data are used for external model validation purpose.

In this study, not only in-sample fitness assessment is conducted with the data set used to construct the model for the purpose of ensuring the model's in-sample fitness. External data forecasting validation is also conducted with two separate data sets along with MAP21 indicators to explore model's forecasting reliability. ND 2012 data set is selected for the purpose of constructing models and ND 2013 and 2014 data sets are selected for external forecasting validation purpose. However, it is easy to demonstrate validation procedures with any data set that makes available.

Bridge distributions by three component ratings for ND 2012 are displayed in Figure 1. As shown in Figure 1, most bridges are coded as 7 or 8 for all superstructure, substructure, and deck conditions. Very few bridges conditions are rated as poor or lower.

In this study, ordinal logistic regression and multiple linear regression models were constructed to forecast bridge conditions. Model fitness and forecasting skills are evaluated and compared between the two types of models. Several criteria were selected for evaluating and comparing the two models and are introduced in the following section.

5. Model Evaluation and Comparison Methods

The following measures were considered in this research: prediction error (PE), bias, accuracy, out-of-range forecasts, percent of correct estimation, Akaike's Information Criteria (AIC), and log likelihood (LL). The models were constructed with the same dataset and compared in two senses: model fitness and prediction performance. All seven proposed measurements can be used to assess model fitness with the same data set that was used to build the model. The first five measures can be used to evaluate model forecasting performance with external evaluation data.

The prediction error, also known as residuals, is a measure of the discrepancy between the observed data and an estimated value which can be mathematically expressed as (3). Two variants of the prediction error were selected for comparison: sum of absolute residuals (SAR) and sum of residual sum of squares (RSS). They are expressed as (4) and (5).

$$e_{(i)} = y_i - \widehat{y}_{(i)} \quad (3)$$

$$SAR = \sum_{i=1}^n |e_{(i)}| \quad (4)$$

$$RSS = \sum_{i=1}^n e_{(i)}^2 \quad (5)$$

where y_i and $\widehat{y}_{(i)}$ are the observed and predicted values of the predictions for data point i .

Bias indicates, on average, how much a model overpredicts (where bias >1) or underpredicts (where bias <1) the

observed data [23], with bias equal to 1 indicating zero bias and it is shown as

$$\text{Bias} = 10^{1/n} \sum_{i=1}^n \log_{10} \left(\frac{\widehat{y}_{(i)}}{y_i} \right) \quad (6)$$

The accuracy measurement indicates, on average, how much the prediction differs from observed data [23], with 1 indicating perfect accuracy. This measurement is shown as

$$\text{Accuracy} = 10^{1/n} \sum_{i=1}^n \left| \log_{10} \left(\frac{\widehat{y}_{(i)}}{y_i} \right) \right| \quad (7)$$

Fit criteria such as Akaike's Information Criteria (AIC) are also selected to compare model fitness between the two models. AIC is a common measure of model fit that balances model fit against model simplicity. The model with the smallest AIC is deemed the "best" model based on apparent validation. In other words, a smaller AIC value indicates a better model/predictor. This can be mathematically expressed as

$$\text{AIC} = n * \ln \left(\frac{\text{RSS}}{n} \right) + 2k \quad (8)$$

where k is the number of free parameters; and n is the number of data points.

Out-of-range forecasts were counted when the forecasted value is greater than 9 and less than 0 for bridge components. This issue only exists for multiple linear regression. For ordinal regression, any out-of-range forecast is always zero. Percent of correct estimation assesses model performance and fitness by examining the prediction and actual observation agreement ratio.

6. Development and Evaluation of Ordinal Logistic Regression Model

Multiple regression models for forecasting bridge component rating are still used by some transportation agencies such as North Dakota DOT to assist in bridge inventory management [8, 16, 23, 24]. In research on this subject, a few key explanatory variables were found to contribute to network bridge rating changes. The variables used in the analysis are summarized in Table 2. The parameters such as weather, freeze-thaw, deicing applications, etc. are not available for us and thus are not included in the model.

Forward stepwise regression based on all adjusted r -square, Akaike information criterion, Bayesian information criterion was used to select the "best" multiple regression model. Detailed regression model selection techniques and theories are out of the scope of this study and readers are referred to Draper and Smith [25]. An ordinal logistic regression model is constructed through an explicit enumeration of all available explanatory variables available and the "best" fitted model is selected based on model selection. The same data and model selection methods are applied for the selection of a multiple linear regression model for the purpose of comparison. In this study, both in-sample fitness

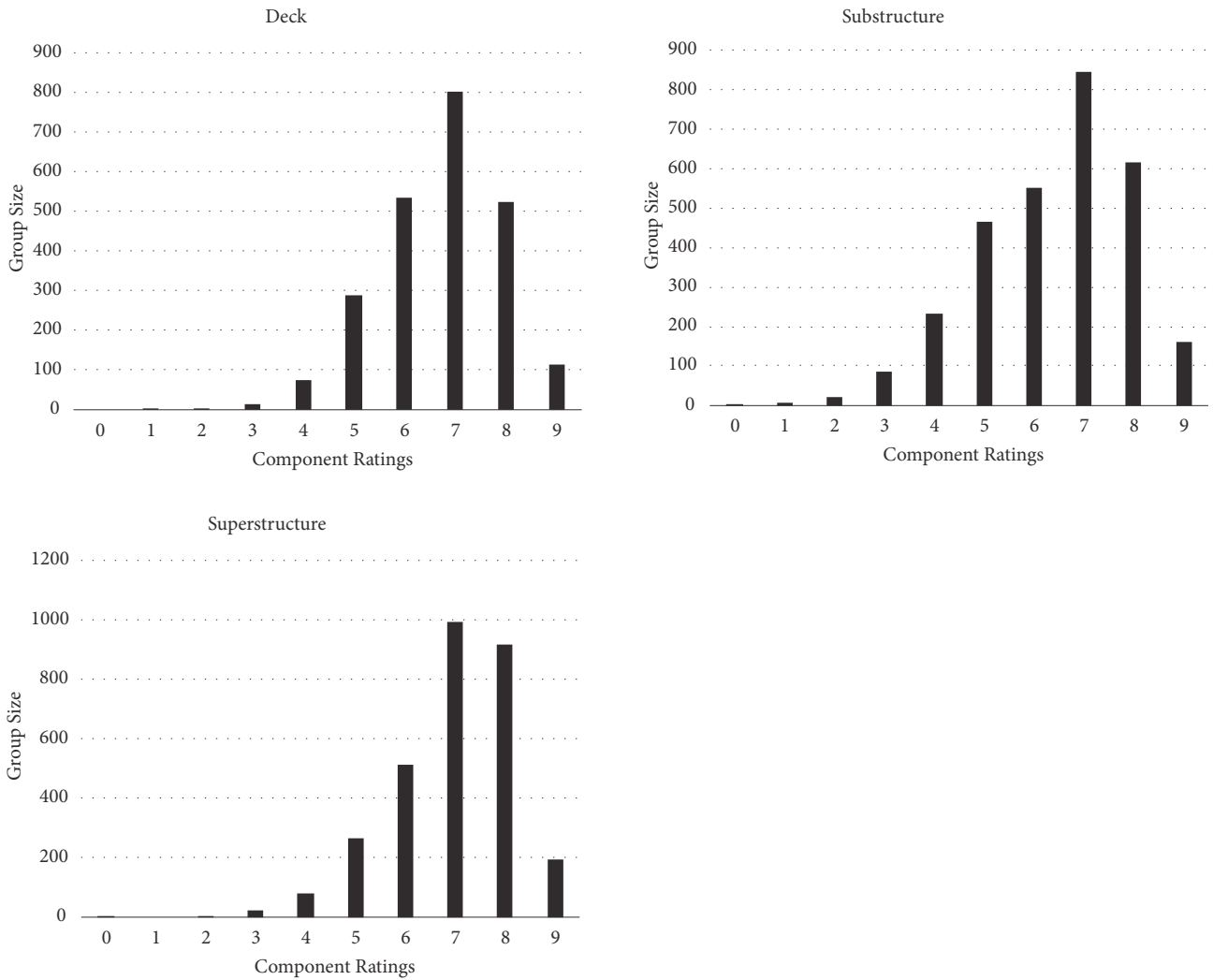


FIGURE 1: Bridge distribution by component ratings for ND 2012.

and external validation will be performed for both regression models and the indicators described in earlier sections will be used to evaluate model quality. In-sample fitness test is conducted with the same data set, ND 2012 data that is used to develop the model. Apparent validation validates a model's goodness of fit on the entire dataset used to construct the model which may not reveal the true predictive ability of the model because the exact same dataset is used for model development and validation test. External validation is the most accurate and unbiased test for the model and the entire data collection process as stated by Harrell et al. [26]. External validation's main emphasis is that predictions from the previously developed model are tested on a new data set that is different from the development population. ND bridge inventory data in 2013 and 2014 can serve on this purpose perfectly for three reasons: (1) NBI reports data annually in separate datasets that will lead to differences in various aspects of the data; (2) the bridges included in validation datasets and model construction datasets vary because bridges may be closed or open to traffic for various

reasons in different years; (3) NBI inspection records in each single year contain the whole bridge population but not sample [27].

Two candidate models were constructed for predicting deck, superstructure, and substructure component performance ratings, respectively, with North Dakota NBI 2012 data. To illustrate the model performance, the models were first evaluated and compared by an in-sample validation method with previously introduced measurements and then ND NBI data from 2013 and 2014 were used to further conduct external prediction validation. Significant parameters for the two sets of models were tested at 90% confidence level as shown in Table 3.

As shown in Table 3, reconstruction, bridge type, and district are categorical variables and ADT, age, and age squared are numerical variables. Positive and negative values in Table 3 indicate the corresponding variables' relationship to component condition categories. For example, ADT has a negative relationship with deck condition. In other words, deck condition ratings decrease as ADT values increase. Both

TABLE 2: Description of variables used in analysis.

Name of variable	Description of variable
Reconstruction	Reconstruction record: Yes, No (binary variable)
Bridge Material Type	Structure materials: Steel, Concrete, Timber (dummy variable)
District	Highway districts: Bismarck, Devils Lake, Dickinson, Grand Forks, Minot, Valley City, Williston, Fargo (dummy Variable)
Age	Bridge age: Inspection year-construction year or inspection year-reconstruction year (continuous variable)
Age2	Bridge age squared (continuous variable)
ADT	Annual daily traffic per lane (continuous variable)

TABLE 3: Significant parameters and statistics with 2012 data.

Model Statistics	Multiple Linear Regression			Ordinal Regression		
	Deck	Superstructure	Substructure	Deck	Superstructure	Substructure
Sample Size	2354	2991	2991	2354	2991	2991
Reconstruction (DF=1)	Significant (Positive for with record comparing without record)	Significant (Positive for with record comparing without record)	Significant (Positive for with record comparing without record)	Significant (Positive for with record comparing without record)	Significant (Positive for with record comparing without record)	Significant (Positive for with record comparing without record)
Bridge Material Type (DF=2)	Significant	Significant	Significant	Significant	Significant	Significant
District (DF=8)	Significant	Significant	Significant	Significant	Significant	Significant
ADT	Significant, Negative	Not Significant, Negative	Significant, Negative	Significant, Negative	Significant, Negative	Significant, Negative
Age	Significant, Negative	Significant, Negative	Significant, Negative	Significant, Negative	Significant, Negative	Significant, Negative
Age2	Significant, Positive	Significant, Positive	Significant, Positive	Significant, Positive	Significant, Positive	Significant, Positive

Note: all independent variables are significant at 90% of the confidence.

age and age squared are identified as significant contributors to condition ratings indicating a nonlinear polynomial effect of age. With positive age squared sign and negative age sign, one can tell the effect of age could be positive up to certain age and ten negative thereafter. For categorical independent variables the relationship is relative among independent variable categories as shown for “reconstruction” which only contains two levels. All models indicate that if a bridge has reconstruction history, the component ratings tend to be better. In other words, reconstruction will improve the bridge component ratings. Note that all models provide expected significant relationships between dependent and independent variables except for ADT’s significance for superstructure rating with the multiple linear regression model. However, the relationship is identified as significant with the ordinal logistic regression model.

To assess how well the model fit the 2012 data, Table 4 shows the comparison results among all the models with indicators introduced earlier in the paper and the percentage of estimations within one or two condition-rating difference. Estimations within 1 condition-rating difference are the estimations that predict component ratings no greater than one above or below observed ratings. Estimations within 2

condition-rating differences are the estimations that predict component ratings no greater than two above or below observed ratings. From Table 4, three ordinal models for deck, superstructure, and substructure consistently show slightly but significant better results than the corresponding multiple linear regression models.

Of the predictions from linear regression models, 0.38%, 0.87%, and 1.07% are out-of-range (0 to 9). The percent of exact-match predictions by three multiple linear regression models (each with the same prediction and observation) are 44.18%, 47.51%, and 41.66%, while the ordinal logistic predictions are much better: 48.3%, 56.74%, and 44.13%. The same conclusion is true for percentage of estimations within one condition-rating difference and within two condition-rating differences. One can tell that the three ordinal multinomial models have more percentage component rating predictions that are off by one or two observation ratings. The bias and accuracy indicators for ordinal logistic regression model are all slightly closer to 1 than those for multiple linear regressions. The sum of absolute residual, sum of residual squares, AIC, and LL, consistently indicate all ordinal regression models perform better than multiple regression models. The ordinal models improve the model performance

TABLE 4: Model comparison statistics with 2012 data.

Model Statistics	Multiple Linear Regression			Ordinal Regression		
	Deck	Superstructure	Substructure	Deck	Superstructure	Substructure
Sample Size	2354	2991	2991	2354	2991	2991
Sum of Absolute Residuals	1,656	1,906	2,380	1,433	1,494	2,200
Sum of Residual Squares	2,446	2,722	4,090	1,933	1,986	3,580
Bias	1.087	1.090	1.128	1.026	1.023	1.078
Accuracy	1.117	1.112	1.180	1.103	1.081	1.174
AIC	5,946	6,875	8,571	5,416	5,994	7,541
LL	-2,957	-3,421	-4,270	-2,845	-3,131	-3,950
Out-of-range Forecasts	0.38%	0.87%	1.07%	0%	0%	0%
Percent Exact Estimations	44%	47%	41%	48 %	56 %	44%
Percent Estimation within 1 Rating Difference	87%	90%	83%	92%	95%	89%
Percent Estimation within 2 Rating Differences	96%	96%	93%	99%	100%	98%

in terms of the sum of residual squares by 20.97%, 27.04%, and 12.47% compared to the multiple regression models for deck, superstructure, and substructure, respectively. Detailed improvement percentage values for all the four indicators are shown in Table 5.

7. Model Validation with New Datasets

To further illustrate the external validation method result, the same ordinal logistic and multiple linear regression models from 2012 data are validated and compared with all ND NBI 2013 and 2014 deck, superstructure, and substructure observed data. Model performances were compared for ordinal logistic and multiple linear regressions by comparing sum of absolute residuals, sum of residual squares, bias, accuracy, out-of-range forecasts, and percentage of estimations which are within one or two rating differences compared with observed component ratings and exact forecasts. The performance results are shown in Tables 6 and 7 for 2013 and 2014, respectively.

One can tell from Tables 6 and 7 that ordinal logistic models perform consistently better than multiple linear regression models for all deck, superstructure, and substructure models and for all indicators with both external validation data sets. The improvement percentages for 2012, 2013, and 2014 are shown in Table 8.

Some interesting observations were obtained in the analysis. Table 8 indicates the superstructure ordinal model has the highest improvement consistently for all three of the yearly data sets and for each of the indicators, except for 2014 bias indicator, followed by the deck model and then the substructure model. The deck model has the highest

improvement for bias in 2014, followed by the superstructure model and then the substructure model. There is no specific trend for model performance improvement by year. For example, deck model improvement for the sum of absolute residuals decreased from 13.47% in 2012 to 13.27% in 2013 and to 12.9% in 2014. The deck model improvement for the sum of residual squares increased from 20.97% to 21.39% then to 22.04% from 2012 to 2014. Finally, the deck model improvement for bias increased from 5.61% to 6.66% and then decreased to 5.7% from 2012 to 2014.

8. Analysis of MAP-21 Bridge Performance Indicator

The MAP-21 rules require all states to report percentage of national highway system bridges classified in good condition and poor condition. Bridge condition can be determined based on an assessment of the deck, superstructure, and substructure. The method used under the Highway Bridge Program is selected to determine bridge conditions: components with condition ratings of no less than 7 are rated as “Good” and no greater than 4 are rated as “Poor”. When all three components are rated as “Good” the overall bridge condition rating can be coded as “Good” and when all three components are rated as “Poor” the overall bridge condition rating can be coded as “Poor”. The observed bridge condition measures and the forecasted measures are listed in Table 9. The values displayed in Table 9 are estimated and observed percentage of bridges in good or poor conditions. By comparing estimated percentage of good or poor bridges to the observed percentages for 2012, 2013, and 2014, respectively,

TABLE 5: Performance improvements by ordinal model compared with multiple linear model.

Model Statistics	Model Performance Improvements Percentage		
	Deck	Superstructure	Substructure
Sum of Absolute Residuals	13.47%	21.62%	7.56%
Sum of Residual Squares	20.97%	27.04%	12.47%
AIC	8.91%	12.81%	12.02%
LL	3.79%	8.48%	7.49%

TABLE 6: Model comparison statistics with 2013 data.

Model Statistics	Multiple Linear Regression			Ordinal Regression		
	Deck	Superstructure	Substructure	Deck	Superstructure	Substructure
Sample Size	2309	2943	2943	2309	2943	2943
Sum of Absolute Residuals	1568	1832	2316	1360	1480	2207
Sum of Residual Squares	2328	2642	4072	1830	1950	3677
Bias	1.096	1.099	1.147	1.023	1.022	1.097
Accuracy	1.126	1.119	1.202	1.099	1.081	1.2
Out-of-range Forecasts	0%	0.2%	0.2%	0%	0%	0%
Percent Exact Forecasts	45%	48%	42%	49%	56%	43 %
Percent Estimation within 1 Rating Difference	88%	90%	84%	93%	94%	87%
Percent Estimation within 2 Rating Differences	98%	97%	94%	100%	100%	97%

TABLE 7: Model comparison statistics with 2014 data.

Model Statistics	Multiple Linear Regression			Ordinal Regression		
	Deck	Superstructure	Substructure	Deck	Superstructure	Substructure
Sample Size	2281	2906	2906	2281	2906	2906
Sum of Absolute Residuals	1550	1793	2247	1350	1441	2129
Sum of Residual Squares	2350	2493	3841	1832	1823	3417
Bias	1.087	1.079	1.105	1.025	1.022	1.053
Accuracy	1.115	1.098	1.156	1.101	1.080	1.151
Out-of-range Forecasts	0%	0%	0.1%	0%	0%	0%
Percent Exact Forecasts	46%	48%	43%	50%	56%	44 %
Percent Estimation within 1 Rating Difference	86%	91%	84%	93%	95%	87%
Percent Estimation within 2 Rating Differences	96%	98%	94%	100%	99%	97%

TABLE 8: Performance improvements by ordinal model compared with multiple linear model.

Model Statistics	2012			2013			2014		
	Deck	Superstruc.	Substruc.	Deck	Superstruc.	Substruc.	Deck	Superstruc.	Substruc.
Sum of Absolute Residuals	13.5%	21.6%	7.56%	13.3%	19.2%	4.71%	12.9%	19.6%	5.25%
Sum of Residual Squares	21.0%	27.0%	12.4%	21.4%	26.2%	9.7%	22.0%	26.8%	11.0 %
Bias	5.61%	6.15%	4.43%	6.66%	7.01%	4.36%	5.7%	5.28%	4.71%
Accuracy	1.25%	2.79%	0.51%	2.4%	3.4%	0.17%	1.26%	1.64%	0.43%
Exact Forecasts	9.33%	19.4%	5.93%	8.01%	15.6%	1.59%	6.46%	16.0%	1.5%

TABLE 9: Bridge condition measures required by MAP-21 comparison results.

	Deck		Superstructure		Substructure		Overall	
	Good	Poor	Good	Poor	Good	Poor	Good	Poor
2012 Observed	48.14%	3.14%	70.38%	3.58%	54.26%	11.84%	44.82%	1.87%
2012 Estimate (Multilinear Regression Model)	60.85%	0.00%	81.71%	0.00%	62.89%	0.90%	54.42%	0.00%
2012 Estimate (Ordinal Model)	51.32%	1.10%	73.65%	0.53%	61.95%	2.67%	48.94%	0.42%
2013 Observed	47.07%	3.11%	69.58%	3.51%	52.56%	12.07%	44.48%	2.04%
2013 Estimate (Multilinear Regression Model)	58.38%	0.00%	80.68%	0.00%	62.39%	2.24%	54.79%	0.00%
2013 Estimate (Ordinal Model)	49.98%	0.13%	73.72%	0.27%	60.22%	5.32%	49.11%	0.13%
2014 Observed	45.14%	3.21%	68.37%	3.18%	51.55%	11.63%	43.05%	1.75%
2014 Estimate (Multilinear Regression Model)	57.14%	0.00%	80.17%	0.00%	61.38%	1.6%	54.49%	0.00%
2014 Estimate (Ordinal Model)	48.85%	0.10%	72.99%	0.50%	59.59%	5.15%	48.27%	0.09%

one can tell that the ordinal model’s estimates are closer to the observed values than multilinear regression estimates.

The above analysis shows that the ordinal logistic models are always better at predicting bridge conditions and measurements with both the in-sample and external validation data sets.

It is worth noting that all the models are underestimate for poor conditions due to the nature of the data distribution. The bridge condition data is imbalanced and biased data set; in other words, the number of observations belonging to one category is significantly lower than those belonging to the other categories. In the situation, the predictive model developed using any GLMs or even conventional machine learning algorithms could be biased and inaccurate. To handle imbalanced classification or improve forecast of rare events data is an extended research and should be investigated in future research.

9. Findings and Discussions

Eight model evaluation criteria were used to compare the goodness of fit and the forecasting power of the models with both in-sample and external validation data sets for deck, superstructure, and substructure condition. The following are the main findings of the study:

- (i) The analysis shows agreement among all indicators, for all three component models, and for all three-year

data sets. All the comparison results indicate the clear improvement of the ordinal logistic model over the multiple linear regression model.

- (ii) Some indicators show significant improvement such as sum of absolute residuals, sum of residual squares, AIC, and exact forecasts (about 10% improvement). However, some indicators show slight improvement such as bias, accuracy, and LL (less than 10% improvement).
- (iii) Superstructure models show the greatest improvement for almost all performance criteria, followed by deck models and substructure models. To further investigate on this issue, time series data need to be tested to confirm that the superstructure model consistently performs better than the other two models
- (iv) There is no clear trend for model performance improvement by year. According to Table 8, for some indicators and certain models, such as sum of residual squares, bias, and accuracy for the deck model, the improvement sequence is 2014, 2013, and 2012 from the greatest improvement to the least improvement by switching from multiple linear regression model to ordinal logistic model. For bias and accuracy of the superstructure model, the sequence is 2013, 2012, and 2014 from the greatest improvement to the least. For sum of residual squares and exact forecasts, the

greatest improvement from switching from a multiple linear regression model to an ordinal logistic model was achieved at year 2012 and then followed by 2014 and 2013.

- (v) Ordinal logistic models will not predict out-of-range estimations which are not controlled by multiple linear regression model.

10. Conclusions

This paper proposes and demonstrates an ordinal logistic regression model for forecasting bridge component rating. The model is preferred for its ability to handle the ordinal nature of bridge component ratings, its explanatory power of the regression analysis, and its accurate prediction power. In this study, both ordinal logistic regression and multiple linear regression models have been generated for predicting three main bridge component ratings. The multinomial logistic model demonstrated in this research can be easily applied with element-level data when it becomes available. In addition to assessing model performance, both in-sample and external validation analysis were performed for all eight evaluation criteria. Finally, it is determined that the ordinal logistic regression method is a better approach than the multiple linear regression method for forecasting bridge component ratings. It has the inherent advantage of always making meaningful predictions and its predictions are closer to the observations.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] D. M. Frangopol, E. S. Gharaibeh, J. S. Kong, and M. Miyake, "Optimal network level bridge maintenance planning based on minimum expected cost," in *Proceedings of the Transportation Research Record*, pp. 26–33, Florida, USA, 2000.
- [2] G. Bu, J. Lee, H. Guan, Y.-C. Loo, and M. Blumenstein, "Long-term performance of bridge elements using integrated deterioration method incorporating elman neural network," *Applied Mechanics and Materials*, vol. 204-208, pp. 1980–1987, 2012.
- [3] O. Thomas and J. Sobanjo, "Comparison of markov chain and semi-markov models for crack deterioration on flexible pavements," *Journal of Infrastructure Systems*, vol. 19, no. 2, pp. 186–195, 2013.
- [4] M. Alsharqawi, T. Zayed, and S. Abu Dabous, "Integrated condition rating and forecasting method for bridge decks using Visual Inspection and Ground Penetrating Radar," *Automation in Construction*, vol. 89, pp. 135–145, 2018.
- [5] M. J. Markow and W. A. Hyman, *Bridge Management Systems for Transportation Agency Decision Making*, NCHRP Synthesis 397, National Academies Press, Washington, DC, USA, 2009.
- [6] D. Tolliver and P. Lu, "Analysis of bridge deterioration rates: a case study of the northern plains region," *Journal of the Transportation Research Forum*, vol. 50, no. 2, 2011.
- [7] S. T. Ariaratnam, A. El-Assaly, and Y. Yang, "Assessment of infrastructure inspection needs using logistic models," *Journal of Infrastructure Systems*, vol. 7, no. 4, pp. 160–165, 2001.
- [8] D. Tolliver and P. Lu, "Modeling bridge condition levels in the United States," *Journal of Civil Engineering and Architecture*, vol. 6, no. 4, pp. 415–432, 2012.
- [9] S. M. Madanat, M. G. Karlaftis, and P. S. McCarthy, "Probabilistic infrastructure deterioration models with panel data," *Journal of Infrastructure Systems*, vol. 3, no. 1, pp. 4–9, 1997.
- [10] Y.-H. Huang, "Artificial neural network model of bridge deterioration," *Journal of Performance of Constructed Facilities*, vol. 24, no. 6, pp. 597–602, 2010.
- [11] J. Ruwanpura, S. T. Ariaratnam, and A. El-Assaly, "Prediction models for sewer infrastructure utilizing rule-based simulation," *Civil Engineering and Environmental Systems*, vol. 21, no. 3, pp. 169–185, 2004.
- [12] O. Thomas and J. Sobanjo, "Semi-Markov models for the deterioration of bridge elements," *Journal of Infrastructure Systems*, vol. 22, no. 3, Article ID 04016010, 2016.
- [13] G. Bu, J. H. Lee, H. Guan, Y. C. Loo, and M. Blumenstein, "Implementation of Elman neural networks for enhancing reliability of integrated bridge deterioration model," *Australian Journal of Structural Engineering*, vol. 15, no. 1, pp. 51–63, 2014.
- [14] G. Nani, I. Mensah, and T. Adjei-Kumi, "Duration estimation model for bridge construction projects in Ghana," *Journal of Engineering, Design and Technology*, vol. 15, no. 6, pp. 754–777, 2017.
- [15] M. Ahmed, O. Moselhi, and A. Bhowmick, "Integration of NDE measurements and current practice in bridge deterioration modeling," in *Proceedings of the 33rd International Symposium on Automation and Robotics in Construction, ISARC 2016*, pp. 341–349, USA, July 2016.
- [16] B. Salman and O. Salem, "Modeling failure of wastewater collection lines using various section-level regression model," *Journal of Infrastructure Systems*, vol. 18, no. 2, pp. 146–154, 2012.
- [17] S. Madanat, R. Mishalani, and W. H. Wan Ibrahim, "Estimation of infrastructure transition probabilities from condition rating data," *Journal of Infrastructure Systems*, vol. 1, no. 2, pp. 120–125, 1995.
- [18] D. Callow, J. Lee, M. Blumenstein, H. Guan, and Y.-C. Loo, "Development of hybrid optimisation method for Artificial Intelligence based bridge deterioration model - Feasibility study," *Automation in Construction*, vol. 31, pp. 83–91, 2013.
- [19] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2002.
- [20] SAS Institute, *PROC LOGISITC Proportional Odds Test and Fitting A Partial Proportional Odds Model*, 2015, <http://support.sas.com/kb/22/954.html>. 2015.
- [21] FHWA, *Information Public Disclosure of National Bridge Inventory (NBI) Data*, 2007, <http://www.fhwa.dot.gov/bridge/nbi/20070517.cfm>.
- [22] FHWA, *Recording and Coding Guide for The Structure Inventory and Appraisal of The Nation's Bridges*, Federal Highway Administration, U.S. Department of Transportation, 1995.

- [23] L. Zhao, Y. Chen, and D. W. Schaffner, "Comparison of logistic regression and linear regression in modeling percentage data," *Applied and Environmental Microbiology*, vol. 67, no. 5, pp. 2129–2135, 2001.
- [24] M. Bolukbasi, J. Mohammadi, and D. Arditi, "Estimating the future condition of highway bridge components using national bridge inventory data," *Practice Periodical on Structural Design and Construction*, vol. 9, no. 1, pp. 16–25, 2004.
- [25] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1981.
- [26] F. E. Harrell Jr., K. L. Lee, and D. B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in Medicine*, vol. 15, no. 4, pp. 361–387, 1996.
- [27] P. Lu, S. L. Pei, and D. Tolliver, "Regression model evaluation for highway bridge component deterioration using national bridge inventory data," *Journal of Transportation Research Forum*, vol. 55, no. 1, pp. 5–16, 2016.

