IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 8, 2017, accepted November 10, 2017, date of publication November 15, 2017,
date of current version December 22, 2017.

*Digital Object Identifier 10.1109/ACCESS.2017.2773612*

# Predicting Crash Rate Using Logistic Quantile Regression With Bounded Outcomes

**XUECAI XU[1,2] AND LI DUAN[1]**
[1]School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan 430073, China
[2]School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Xuecai Xu (xuecai_xu@hust.edu.cn)

This work was supported by the Fundamental Research Fund for the Central Universities under Grant HUST: 2016YXM099 and Grant 2016YXM100.

**ABSTRACT** Various approaches and perspectives have been presented in safety analysis during the last decade, but when some continuous outcome variables take on values within a bounded interval, the conventional statistical methods may be inadequate, and frequency distributions of bounded outcomes cannot be used to handle it appropriately. Therefore, in this paper, a logistic quantile regression (QR) model is provided to fill this gap and deal with continuous bounded outcomes with crash rate prediction. The crash data set from 2003 to 2005 maintained by the Nevada Department of Transportation is employed to illustrate the performance of the proposed model. The results show that average travel speed, signal spacing, driveway density, and annual average daily traffic on each lane are significantly influencing factors on crash rate, and logistic QR is verified as an alternative method in predicting crash rate.

**INDEX TERMS** Crash rate, logistics quantile regression, bounded outcomes.

## I. INTRODUCTION

Access management techniques can be used to control the location, spacing design and operation of driveways, median treatments, median openings, and related auxiliary lanes systematically. The goal of access management is to perform the roadway functional hierarchy in the transportation planning, which provides the accessibility to the adjacent land as well as the safety and efficiency of the transportation system (particularly along arterials and other primary roadways). Thus, identification of access management factors, such as signal spacing, driveway spacing, median treatments, etc., is required to support safe and efficient operations for the main arterials, as well as providing the convenient accesses for the adjacent land.

During the last decade, there have been a variety of different approaches and perspectives [1]–[3] presented in safety evaluation, and the heterogeneity issue has been addressed by various studies through finite mixture regression models [4] and random parameter model [5], [6], in which the heterogeneity from the data or locations caused by unobserved factors was accommodated, and the estimation results and statistical inferences were improved [7]–[12]. However, crashes occurring at the same arterial probably share common unobserved factors. The distributional assumption required to estimate the random parameters may not adequately address this unobserved feature [3]. Most importantly, the models mentioned above all belong to mean regression, in which the model assumptions cannot be easily extended to non-central distributions and are not always suitable with real-world data, especially in the case of homoscedasticity [13], thus a more appropriate and more complete view is required to capture the distributional properties with a broader spectrum than only mean and variance. Inspired by the statements above, the purpose of this study is to introduce an innovative model in road safety analysis to identify the influencing factors of access management in urban arterials.

In recent years, quantile regression (QR), initially proposed by Koenker and Bassett (1978) [14], has attracted increasing attention in various fields, e.g. sociology, economics, finance and medical science [13], [15]. Quantile regression is a powerful tool for comparing, more thoroughly than the mean alone, various aspects (location, scale, and shape) of any kind of distribution of the outcome across different covariate patterns [16]. The main merit of quantile regression is that quantile regression does not require the data to follow a specific distribution, but estimate multiple variations from several regression curves for different percentage points of the distribution, which may reflect different effects at different quantiles of the response variable. Moreover, quantile regression is more robust against outliers because the estimation results may be less sensitive to outliers and multi-modality [17]. In particular, quantile regression can handle the heterogeneity issue for the data collected from different sources at different locations and different times without many assumptions [13], [18], [19], which is helpful to describe the relationship between safety and access management factors more clearly.

With thirty years' progress, QR has been used in various fields and areas, whereas the application in transportation field remains sparse [13], [18]–[21]. The pioneering study by Hewson (2008) [22] examined the potential role of quantile regression for modeling the speed data, and demonstrated the potential benefits of using quantile regression methods, providing more interest than the conditional mean. From the perspective of discrete variables, Qin *et al.* (2010) [18] identified crash-prone locations with quantile regression. The flexibility of estimating trends at different quantiles was offered, and the data with heterogeneity was tackled. The findings suggest that quantile regression yields a sensible and much more refined subset of risk-prone locations. Next, Qin and Reyes (2011) [19] and Qin (2012) [13] modeled crash frequencies with quantile regression. QR tackles heterogeneous crash data and offers a complete view of how the covariates affect the responsible variable from the full range of the distribution, which is beneficial for data with heavy tails, heteroscedasticity and multi-modality. The results show that quantile regression estimates are more informative than conditional means. Similarly, Wu *et al.* (2014) [20] analyzed crash data using quantile regression for counts. The results revealed more detailed information on the marginal effect of covariates change across the conditional distribution of the response variable, and provided more robust and accurate predictions on crash counts. After that, Liu *et al.* (2013) [17] analyzed the train derailment severity using zero-truncated negative binomial regression and quantile regression, and provided insights for train derailment severity under various operational conditions and by different accident causes. From the perspective of identifying accident blackspots in a transportation network, Washington *et al.* (2014) [21] applied quantile regression to model equivalent property damage only (PDO) crashes. The proposed method identified covariate effects on various quantiles of the population and performed better than traditional Negative Binomial (NB) model.

Moreover, the studies of QR have been extended and applied to other areas. Kniesner *et al.* (2010) [23] examined differences in the value of statistical life (VSL) across potential wage levels using panel data quantile regressions with intercept heterogeneity. The findings indicated that VSL varies considerably across the labor force. Delisi *et al.* (2011) [24] revisited the criminal specialization with simultaneous quantile regression. The results implied that age, sex and arrest onset had differential predictive validity of specialization at different quantiles. Recently, Arunraj and Ahrens (2015) [25] combined seasonal autoregressive integrated moving average (SARIMA) with quantile regression for daily food sales forecasting. The results showed that SARIMA-QR model yielded better forecasts at out-sample data and provided a deep insight into the effects of demand influencing factors for different quantiles compared to traditional SARIMA.

However, when some continuous outcome variables take on values within a bounded interval, the conventional statistical methods, such as least squares regression, mixed-effects models, may be inadequate, and frequentist methods of bounded outcomes cannot be used to handle it appropriately. Literally, logistic quantile regression model can provide an effective method to fill this gap and deal with continuous bounded outcomes with consequences. In accordance with this, Bottai *et al.* (2010) [26] firstly described the use logistic quantile regression with bounded outcome in the biomedical and epidemiological areas. Recently, it was extended to other areas: Feizi *et al.* (2012) [27] investigated the association between perceived stress and major life events stressors in Iranian general population. Logistic quantile regression was used for modeling as the bounded outcome, and the results showed that family conflicts and social problems were more correlated with level of perceived stress. Siao *et al.* (2016) [28] predicted recovery rates using logistic quantile regression with bounded outcomes and the empirical results confirmed that the method was more robust in the accuracy performance. Notably, there is no application of logistic quantile regression model in transportation field, thus this study would be the initial attempt to employ logistic quantile regression model to predict crash rate.

## II. METHODOLOGY

Consider a sample of n observations on continuous outcome variable $y_i, i = 1, \ldots n$, and a set of covariates $x_i = \{x_1, x_2, \ldots x_s\}$, where $y_i$ is bounded within a known interval $y_{\min}$ and $y_{\max}$. Here $y_{\min}$ and $y_{\max}$. do not denote the smallest and the largest observed sample values, but the limits of the feasible interval of the outcome variable. The quantile regression model can be expressed as

$$y_i = x_i\beta_p + \varepsilon_i \tag{1}$$

where the $\beta_p = \{\beta_{p1}, \beta_{p2}, \ldots, \beta_{ps}\}$ represents the unknown regression parameters. Let p be a number between zero and one, assume $P(\varepsilon_i \leq 0\,|x_i) = p$ or equivalently $P(y_i \leq x_i\beta_p\,|x_i) = p$. The p quantile of the conditional distribution of $y_i$ given $x_i$ can be described as

$$Q_y(p) = x_i\beta_p \tag{2}$$

If $p = 0.5$, $Q_y(0.5)$ is the conditional median, the value that splits the conditional distribution of the outcome variable into two parts with equal probability. Besides this, the regression residual $\varepsilon_i$ does not require other assumptions.

Contrary to the mean regression, QR estimation is invariant to monotonic transformations of the outcome, that is, $Q_{h(y)}(p) = h\{Q_y(p)\}$ for any non-decreasing function h, while $E\{h(y)\} \neq h\{E(y)\}$ where $E(y)$ represents the mean of $y$. It is assumed that for any quantile p there exists a fixed set of parameters $\beta_p$ and the non-decreasing function h from the interval $(y_{\min}, y_{\max})$ to the real line (often referred to as a link), such that

$$h\{Q_y(p)\} = x_i\beta_p \quad (3)$$

Because a continuous outcome bounded within the unit interval resembles a probability, similar to a most popular regression method for binary outcomes, among a variety of suitable options for the link function h, here the logistic transformation is selected:

$$h(y_i) = \log(\frac{y_i - y_{\min}}{y_{\max} - y_i}) = \log it(y_i) \quad (4)$$

By integrating Equation (3) and Equation (4), the inverse function can be expressed as follows:

$$Q_y(p) = \frac{\exp(x_i\beta_p)y_{\max} + y_{\min}}{1 + \exp(x_i\beta_p)} \quad (5)$$

By regression the transformed outcome $h(y_i)$ on x, the estimated regression coefficients can be achieved via quantile regression. When the estimation for the regression coefficients $\beta_p$ is obtained, inference on $Q_y(p)$ can be made through the inverse transform in Equation (5),

$$Q_{h(y_i)}(p) = Q_{\log it(y_i)}(p) = x_i\beta_p \quad (6)$$

which is analogous to logistic regression, utilizing the same transform to model a probability. Regarding inference about $\beta_p$, it has been indicated that in quantile regression bootstrap standard errors outperform asymptotic ones [16]. Therefore, in Stata software bootstrap is considered as the default method for estimating the standard errors. More details can be referred to [16] and [26].

## III. DATA DESCRIPTION

The crash geo-dataset from 2003 to 2005 was collected from Nevada Department of Transportation, while the access management and roadway characteristics, annual average daily traffic (AADT) and geometric features were integrated from Google Earth correspondingly. The target population is located in Las Vegas metropolitan area. A total of 400 roadway segments with 27 major and minor arterials were sampled as shown in Fig.1. QR serves as the desirable option for modeling crash rate (expressed as crashes per million vehicles miles travelled (MVMT)) because the distribution of crash rate is skewed as shown in Fig. 2. The way that crash rate is considered is because incorporates the effect of volume and segment length, it is more adequate to measure the crash risk faced and perceived by individual drivers than crash frequency, which is highly related to the traffic volume. Another reason is that QR model requires the dependent variable be continuous, which can't be replaced by crash frequency because it's discrete. Furthermore, due to the data collection process, substantial heterogeneity exists in the crash data.

The major variables included are mostly related to access management techniques, such as signal spacing, driveway density, median types, median opening density, etc. Moreover, the roadway characteristics and other features are also involved, e.g. AADT, number of lanes, and land use types. Table 1 summarizes the statistics of main variables. More detailed information on data collection and processing can be found in [29] and [30].

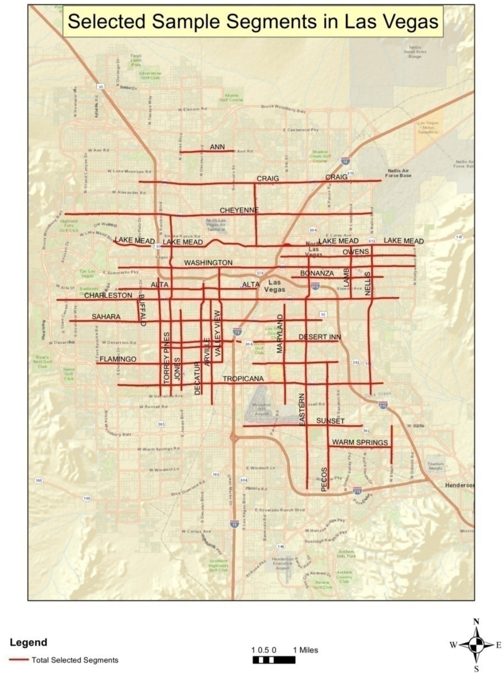



FIGURE 1. Selected segment in las vegas.

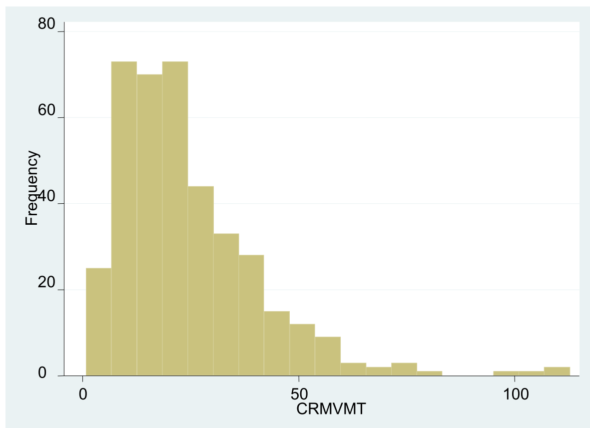



FIGURE 2. Histogram of crash rate.

## IV. RESULTS AND DISCUSSION

As stated in the modeling, bootstrap estimation method is employed to predict crash rate, and confidence intervals are calculated for each estimated coefficient with STATA 14. After all the variables were input into the model, the final form is as follows:

$$Q_{\log it(CRMVMT)}(p) = \beta_{p,0} + \beta_{p,1}AVGSP + \beta_{p,2}SGSPACING + \beta_{p,3}DWDEN + \beta_{p,4}AADTLN$$

**TABLE 1.** Summary statistics for main variables.

| Variable | Description | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| **CRMVMT** | Crashes per MVMT | 24.50 | 16.82 | 0.8 | 112.74 |
| **AVGSP** | Average travel speed, mph | 38.41 | 4.37 | 20.36 | 43.00 |
| **SGSPACING** | Signal spacing, mile | 0.42 | 0.19 | 0.06 | 1.05 |
| **DWDEN** | Driveway density, number of driveways per mile | 34.64 | 19.18 | 0 | 95.78 |
| **MEDOPDEN** | Total median opening density, total number of median openings per mile | 2.89 | 3.28 | 0 | 13.89 |
| **TWODIRDEN** | Two-directional median opening density, number of two-directional median openings per mile | 0.94 | 1.76 | 0 | 8.21 |
| **ONEDIRDEN** | One-directional median opening density, number of one-directional median openings per mile | 1.42 | 2.54 | 0 | 15.61 |
| **RESDEN** | Residential land density, number of residential lands per mile | 8.00 | 7.10 | 0 | 31.52 |
| **COMDEN** | Commercial land density, number of commercial lands per mile | 24.47 | 16.47 | 0 | 87.22 |
| **POSTSP** | Posted speed limit, mph | 41.76 | 4.84 | 30 | 45 |
| **AADTLN** | Annual average daily traffic per thousand per lane | 7.49 | 2.87 | 0.25 | 14.84 |
| | | **Frequency** | | **Proportion (%)** | |
| **MEDTYP** | 1 is for raised median | 224 | | 56.7 | |
| | 0 for two-way-left-lane (TWLTL) | 171 | | 43.3 | |

**TABLE 2.** Estimation results for logistic quantile and logistic regression models.

| Variable | Quantile | | | | | Logistic |
|---|---|---|---|---|---|---|
| | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | **Mean** |
| **AVGSP** | -0.092* (-7.18) | -0.081* (-7.04) | -0.077* (-6.25) | -0.073* (-4.34) | -0.061 (-1.52) | -0.064* (-7.66) |
| **SGSPACING** | -0.627* (-2.77) | -1.176* (-4.06) | -1.413* (-4.39) | -1.165* (-2.97) | -1.552* (-3.14) | -0.827* (-5.17) |
| **DWDEN** | 0.017* (7.92) | 0.014* (6.56) | 0.013* (3.93) | 0.011* (3.12) | 0.009 (1.40) | 0.011* (7.16) |
| **AADTLN** | -0.174* (-6.73) | -0.197* (-9.22) | -0.200* (-10.21) | -0.209* (-8.41) | -0.209* (-4.79) | -0.139* (-11.59) |
| **Constant** | 2.556* (3.51) | 3.085* (5.64) | 3.356* (6.73) | 3.996* (4.51) | 3.845* (2.02) | 6.446* (16.88) |

Note:* indicates significant at the 5% level.

The model was firstly estimated with all the variables and then eliminated the most insignificant ones step by step at the 5% level of significance until all of the variables remaining were significant. Table 2 reports the estimated coefficients and 95% confidence intervals for statistically significant variables at the 25th, 50th, 75th, 90th, and 95th percentile of crash rate distribution. Consequently, it presents a broader and complete view of the variables with different crash rates, that is to say, rather than assuming the coefficient are fixed across all the arterials, some or all of them are allowed to vary to account for heterogeneity attributed to unobserved factors.

The variables total median opening density, two-directional median opening density, residential land density, posted speed limit and median type are neglected from the table because they are not statistically significant, while the variable commercial land density is omitted because it is highly correlated with driveway density from correlation test.

As shown from Table 2, although the logistic regression and logistic quantile regression model share the same significant variables, a closer examination of the magnitude of the estimated coefficients reveals some similarities and differences among quantiles. First, all the covariates in the final model appear to influence the crash rate, and they are all significant in nearly every quantile considered. With only the exception of the 0.95 quantile, there is increasing trend of similar magnitudes for average travel speed percentiles, while there is decreasing trend for the other three significant variables.

Secondly, AADT on each lane and signal spacing are two of the most important variables, which are of significance from the 0.25 to the 0.95 quantile. But the average travel speed and driveway density are not statistically significant at the 0.95 quantile. As shown from Figure 1, the distribution of crash rate concentrates before 90%; correspondingly, the impact of driveway density and travel speed on crash rate begins to weaken till the 95th percentile. The reason that the two variables are not statistically significant at the 95th percentile is not only due to the shortage of crash data, but also because of other influencing factors besides the listed variables, such as vehicle problems, drivers' issues, and even environmental conditions, etc. This suggests that the diversity of data sets may need to be considered when evaluating the safety impact at the mid-block segments.

Fig. 3 gives the estimates (solid lines) and the 95 percent confidence bands (shaded gray areas) for the regression coefficients associated with the significant variables for a dense set of quantiles. The horizontal line at zero is marked for reference. The figure depicts the information shown numerically in Table 2 for five quantiles and extends it to a larger set of quantiles. The confidence bands allow visual inspection of the import of the sampling error.

Corresponding to Table 2, the coefficient of average travel speed increases as p increase, and is significant for all quantiles except 0.95, with confidence band far below zero as shown in Figure 3(a). When average travel speed increases by one unit, the 0.25 quantile of the logit of crash rate is

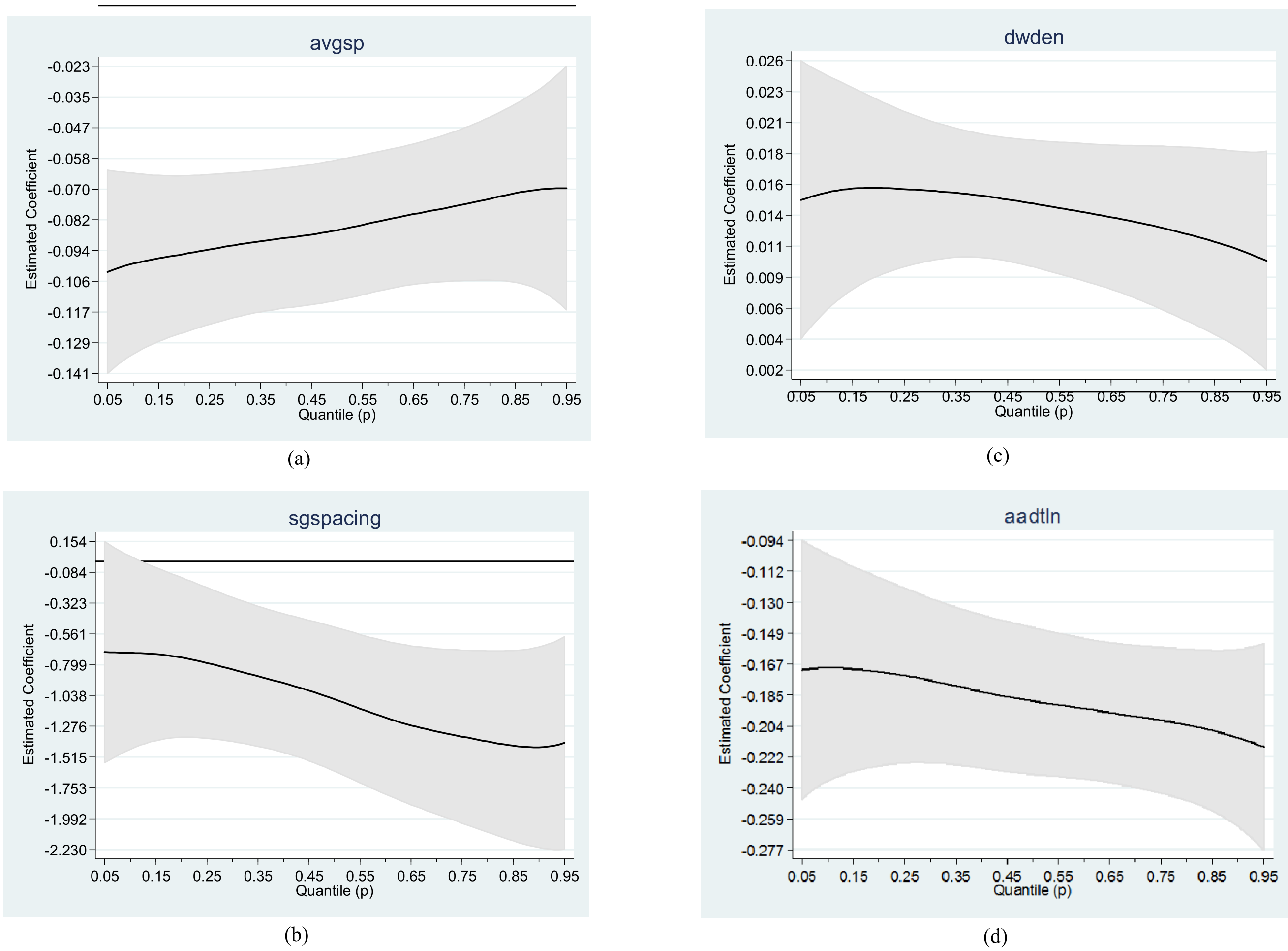



**FIGURE 3. Estimates for the regression coefficients.**

estimated to decrease by 0.092 units whereas the 0.90 quantile by 0.073 only, which can be inferred that the beneficial effect of average travel speed is stronger in the lower quantile of the population than it is in the upper. Therefore, the average travel speed is more influential in lower quantile than in those with high quantiles, though the whole trend is kept increasing.

As shown in Figure 3(b), the effect of signal spacing on crash rate is downward in the total trend, yet still significant in all quantiles. It can be inferred that longer signal spacing reduces the crash rate. However, from 90th quantile the confidence band starts to increase a little, which implies that the crash rate would not be reduced when the signal spacing reaches certain limit.

The driveway density is significantly different from zero for all quantiles except 95th quantile. It can be seen from Figure 3(c) that before 25th quantile the effect of driveway density is upward, meaning that the larger the driveway density the higher crash rate, but the trend turns to downward beyond 25th quantile, although the coefficients at the 50th quantile (estimate 0.014) and at the 90th quantile (estimate 0.011) yields a difference less than 0.01.

The AADT on each lane is significant in all quantiles considered, but the effect trend on crash rate is upward before 0.15 quantile and downward after that, implying that the larger the AADT on each lane, the higher the crash rate before 0.15 quantile whereas the lower the crash rate after 0.15 quantile. Usually if there is more traffic on each lane, the chances of running into the crashes are higher, while when the AADT per lane is large enough, the travel speed would be lower, thus the possibility of running into conflicts may be lower.

In order to compare the errors of logistic quantile and logistic regression models, root-mean-square-error (RMSE) is employed to measure the difference between values predicted and observed in the field. The RMSE is defined as $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y})^2}$, where $Y_i$ is the observed value, $\hat{Y}_i$ is the predicted value and $n$ is the number of observations. Table 3 gives the RMSEs of two models. The results show that on average RMSE values of the logistic quantile regression models are smaller than that of the logistic regression model in this study, especially smaller at 0.5 quantile.

**TABLE 3.** Comparison of RMSES for logistic quantile and logistic regression model.

| Model | Logistic quantile | | | | | Logistic |
|---|---|---|---|---|---|---|
| Quantile | 0.25 | 0.5 | 0.75 | 0.90 | 0.95 | |
| RMSE | 0.477 | 0.437 | 0.439 | 0.479 | 0.497 | 0.544 |

## V. CONCLUSIONS

In this study the quantile regression with a logistic transform was explored for the crash rate analysis of continuous outcomes bounded from above and below by known values. The crash dataset from 2003 to 2005 maintained by Nevada Department of Transportation was employed to illustrate the performance of proposed model. The results show that average travel speed, signal spacing, driveway density and AADT on each lane are significantly influencing factors on crash rate.

Two critical findings can be concluded from the results of the study. First, Logistic quantile regression can be considered as a practical approach to inference about the conditional distribution of bounded outcomes given a set of covariates. Its inference is valid with any underlying distribution, and it allows a deeper understanding than the mean regression methods. Second, the confidence bands lie in its entirety, instead of the mean. The estimates of all the regression coefficients reflect the whole trend variation, and reveal a complete spectrum, which gives a full view of the significant variables.

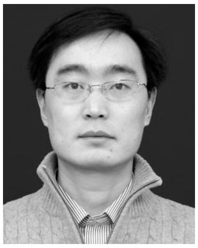
However, one deficiency of this study is that the data collected were mostly related to access management techniques, and other parameters (e.g. weather conditions, signal phases, and other attributes) may still play some role in the impact of crash rate, which may influence the accuracy of the propose model, and need to be supplemented in the future.

## ACKNOWLEDGMENT

The authors are grateful to the Nevada Department of Transportation for providing the database used in this research.

## REFERENCES

[1] D. Lord and F. L. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 291–305, Jun. 2010.
[2] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives," *Accid. Anal. Prev.*, vol. 43, no. 5, pp. 1666–1676, Sep. 2011.
[3] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Anal. Methods Accident Res.*, vol. 1, pp. 1–22, Jan. 2014.
[4] E. S. Park and D. Lord, "Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2019, pp. 1–6, Feb. 2007.
[5] P. C. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle accident frequencies with random-parameters count models," *Accident Anal. Prevention*, vol. 41, no. 1, pp. 153–159, Jan. 2009.
[6] K. El-Basyouny and T. Sayed, "Accident prediction models with random corridor parameters," *Accident Anal. Prevention*, vol. 41, no. 5, pp. 1118–1123, Sep. 2009.
[7] P. C. Anastasopoulos and F. L. Mannering, "An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data," *Accident Anal. Prevention*, vol. 43, no. 3, pp. 1140–1147, May 2011.
[8] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Anal. Prevention*, vol. 45, pp. 628–633, Mar. 2012.
[9] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Anal. Methods Accident Res.*, vol. 11, pp. 1–16, Sep. 2016.
[10] P. C. Anastasopoulos, "Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis," *Anal. Methods Accident Res.*, vol. 11, pp. 17–32, Sep. 2016.
[11] Q. Zeng, H. Wen, H. Huang, and M. Abdel-Aty, "A Bayesian spatial random parameters Tobit model for analyzing crash rates on roadway segments," *Accident Anal. Prevention*, vol. 100, pp. 37–43, Mar. 2017.
[12] G. Fountas and P. C. Anastasopoulos, "A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities," *Anal. Methods Accident Res.*, vol. 15, pp. 1–16, Sep. 2017.
[13] X. Qin, "Quantile effects of casual factors on crash distributions," *Transp. Res. Rec.,J. Transp. Res. Board*, vol. 2279, pp. 40–46, 2012, doi: 10.3141/2279-05.
[14] R. Koenker and G. Bassett, Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, Jan. 1978.
[15] Y. Wang, X.-N. Feng, and X.-Y. Song, "Bayesian quantile structural equation models," *Struct. Equ. Model., Multidisciplinary J.*, vol. 23, no. 2, pp. 246–258, Jul. 2015.
[16] N. Orsini and M. Bottai, "Logistic quantile regression in Stata," *Stata J.*, vol. 11, no. 3, pp. 327–344, 2011.
[17] X. Liu, M. R. Saat, X. Qin, and C. P. L. Barkan, "Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression," *Accident Anal. Prevention*, vol. 59, pp. 87–93, Oct. 2013.
[18] X. Qin, M. Ng, and P. E. Reyes, "Identifying crash-prone locations with quantile regression," *Accident Anal. Prevention*, vol. 42, no. 6, pp. 1531–1537, Nov. 2010.
[19] X. Qin and P. E. Reyes, "Conditional quantile analysis for crash count data," *J. Transp. Eng.*, vol. 137, no. 9, pp. 601–607, Sep. 2011.
[20] H. Wu, L. Gao, and Z. Zhang, "Analysis of crash data using quantile regression for counts," *J. Transp. Eng.*, vol. 140, no. 4, p. 04013025, Apr. 2014, doi: 10.1061/(ASCE)TE.1943-5436.0000650.
[21] S. Washington, M. M. Haque, J. Oh, and D. Lee, "Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots," *Accident Anal. Prevention*, vol. 66, pp. 136–146, May 2014.
[22] P. Hewson, "Quantile regression provides a fuller analysis of speed data," *Accident Anal. Prevention*, vol. 40, no. 2, pp. 502–510, Mar. 2008.
[23] T. J. Kniesner, W. K. Viscusi, and J. P. Ziliak, "Policy relevant heterogeneity in the value of statistical life: New evidence from panel data quantile regressions," *J. Risk Uncertainty*, vol. 40, no. 1, pp. 15–31, Feb. 2010.
[24] M. Delisi, K. M. Beaver, K. A. Wright, J. P. Wright, M. G. Vaughn, and C. R. Trulson, "Criminal specialization revisited: A simultaneous quantile regression approach," *Amer. J. Criminal Justice*, vol. 36, no. 2, pp. 73–92, Jun. 2011.
[25] N. S. Arunraj and D. Ahrens, "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," *Int. J. Prod. Econ.*, vol. 170, pp. 321–335, Dec. 2015.
[26] M. Bottai, B. Cai, and R. E. McKeown, "Logistic quantile regression for bounded outcomes," *Stat. Med.*, vol. 29, no. 2, pp. 309–317, Nov. 2009.
[27] A. Feizi, R. Aliyari, and H. Roohafza, "Association on perceived stress with stressful life events, lifestyle and sociodemographic factors: A large-scale community-based study using logistic quantile regression," *Comput. Math. Methods Med.*, vol. 2012, Sep. 2012, doi: 10.1155/2012/151865.
[28] J.-S. Siao, R.-C. Hwang, and C.-K. Chu, "Predicting recovery rates using logistic quantile regression with bounded outcomes," *Quant. Finance*, vol. 16, no. 5, pp. 777–792, Jul. 2015.
[29] X. Xu, V. Kwigizile, and H. Teng, "Identifying access management factors associated with safety of urban arterials mid-blocks: A panel data simultaneous equation models approach," *Traffic Injury Prevention*, vol. 14, no. 7, pp. 734–742, 2013.
[30] X. Xu and A. Kouhpaenejad, "Simultaneous equation Tobit analysis of arterial road crash rates," in *Proc. Inst. Civil Eng.-Transp.*, Aug. 2015, vol. 168. no. 4, pp. 346–355.



**XUECAI XU** received the B.S. degree from the Xi'an University of Technology, Xi'an, China, the M.E. degree from Southwest Jiaotong University, Chengdu, China, and the Ph.D. degree from the University of Nevada, Las Vegas, NV, USA.

He was a Senior Research Assistant with the Department of Civil Engineering, The University of Hong Kong, Hong Kong. He is a Lecturer with the School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China. He is currently a Senior Research Fellow with the School of Civil and Environmental Engineering, Nanyang Technological University, Singapore.

He has been a PI or Co-PI of various funds, including the Fundamental Research Fund for the Central Universities, the Scientific Research Foundation for Returned Overseas Chinese Scholars, the State Education Ministry of China, and the National Natural Science Foundation of China. He regularly involved in projects with industry and government agencies in China and Overseas. He has authored over 20 articles indexed by SCI/SSCI/EI. His research interests include transportation safety, intelligent transportation systems, and system engineering.

Dr. Xu was a member of the Institute of Transportation Engineering. He serves as a Reviewer for the *Accident Analysis and Prevention*, the *International Journal of Sustainable Transportation*, the *Journal of Advanced Transportation*, the *Traffic Injury Prevention*, the *Transportmetrica*, the *Mathematical Problems in Engineering*, and the *Discrete Dynamics in Nature and Society*.

**LI DUAN** received the bachelor's and master's degrees from Southwest Jiaotong University, Chengdu, China, and the Ph.D. degree from Southeast University, Nanjing, China.

He is currently a Lecturer with the School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China. His research interest includes transportation safety, intelligent transportation systems, and transportation engineering analysis.