Accident Analysis and Prevention 59 (2013) 87–93

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression

Xiang Liu[a,*], M. Rapik Saat[a,1], Xiao Qin[b,2], Christopher P.L. Barkan[a,3]

[a] *Rail Transportation & Engineering Center (RailTEC), University of Illinois at Urbana-Champaign, Newmark Civil Engineering Laboratory, 205 North Mathews Avenue, Urbana, IL 61801, United States*

[b] *Department of Civil and Environmental Engineering, South Dakota State University, CEH 148, Box 2219, Brookings, SD 57007, United States*

ARTICLE INFO

Article history:
Received 22 January 2013
Received in revised form 24 April 2013
Accepted 29 April 2013

Keywords:
Rail
Derailment severity
Zero-truncated negative binomial
Quantile regression

ABSTRACT

Derailments are the most common type of freight-train accidents in the United States. Derailments cause damage to infrastructure and rolling stock, disrupt services, and may cause casualties and harm the environment. Accordingly, derailment analysis and prevention has long been a high priority in the rail industry and government. Despite the low probability of a train derailment, the potential for severe consequences justify the need to better understand the factors influencing train derailment severity. In this paper, a zero-truncated negative binomial (ZTNB) regression model is developed to estimate the conditional mean of train derailment severity. Recognizing that the mean is not the only statistic describing data distribution, a quantile regression (QR) model is also developed to estimate derailment severity at different quantiles. The two regression models together provide a better understanding of train derailment severity distribution. Results of this work can be used to estimate train derailment severity under various operational conditions and by different accident causes. This research is intended to provide insights regarding development of cost-efficient train safety policies.

1. Introduction

Railways are vital for U.S. national economy. While the society derives significant benefits from rail transportation, there are certain safety risks that must be managed and minimized to a feasible extent. Train accidents cause damage to infrastructure and rolling stock, disrupt services, and may cause casualties and harm the environment. Derailments accounted for 72% of freight-train accidents in the United States from 2001 to 2010 (Liu et al., 2012). Correspondingly, the analysis and prevention of train derailments is a high priority in the rail industry and government.

The probability of a train derailment has been studied by previous researchers (e.g., Nayak et al., 1983; Treichel and Barkan, 1993; Dennis, 2002; Anderson and Barkan, 2004; Kawprasert, 2010; Liu et al., 2011, 2012). In addition to analyzing the likelihood of a derailment, understanding the magnitude and variability of derailment severity is equally important. In this paper, derailment severity is measured by number of cars derailed after a train derailment occurs. The generic use of "cars" refers to all vehicles (including locomotives, railcars and cabooses), unless specifically stated otherwise. Quantifying the relationship between train derailment severity and associated affecting factors could aid the rail industry and government to develop, evaluate, prioritize and implement cost-effective safety improvement strategies.

2. Literature review

Simulation and statistical analysis are the two basic approaches used in previous studies to model train derailment severity. Simulation models predict the response of railroad vehicles to specific track and environmental conditions. These models are typically based on detailed nonlinear wheel-rail interaction models. For example, Yang et al. (1972, 1973) developed a simulation model to determine the effect of ground friction, mating coupler moment, and brake retarding force on the number of cars derailed. They found that the position of the first car involved in the derailment (called point-of-derailment, or POD) and derailment speed could affect the number of cars derailed (Yang et al., 1972, 1973). In the late 1980s, Yang et al.'s model was extended by considering coupler failure and independent car motion (Coppens et al., 1988; Birk et al., 1990). The precision of simulation models is subject to the accuracy of modeling train derailment dynamics.

---

* Corresponding author. Tel.: +1 217 244 6063.
E-mail addresses: liu94@illinois.edu (X. Liu), mohdsaat@illinois.edu (M.R. Saat), xiao.qin@sdstate.edu (X. Qin), cbarkan@illinois.edu (C.P.L. Barkan).

[1] Tel.: +1 217 333 6974.

[2] Tel.: +1 605 688 6355.

[3] Tel.: +1 217 244 6338.

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.04.039

In addition to simulation models, train derailment severity can also be estimated based on historical data. Saccomanno and El-Hage (1989, 1991) developed a truncated geometric model to estimate the mean number of cars derailed as a function of derailment speed, residual train length and accident cause. The model was modified by Anderson (2005) and Bagheri (2009), respectively. There is interest to consider new factors that may affect derailment severity. Last but not least, all previous derailment severity models focused on analyzing the mean number of cars derailed. Depending on factors discussed later in this paper, other distributional statistics may also need to be understood, such as quantiles.

3. Train derailment severity

The number of cars derailed in freight-train derailments on U.S. Class I[4] railroad mainlines, from 2001 to 2010, is plotted in Fig. 1. On average, a freight-train derailment resulted in approximately 10 cars derailed, and the median severity is 6 cars derailed.

The literature has investigated the effect of accident cause on train derailment severity (e.g., Saccomanno and El-Hage, 1989, 1991; Barkan et al., 2003; Anderson, 2005; Bagheri, 2009; Bagheri et al., 2011; Liu et al., 2011, 2012). Broken rails are the most common cause of freight-train derailments on U.S. Class I mainlines (Barkan et al., 2003; Liu et al., 2012). On average, a broken-rail-caused freight-train derailment caused 14 cars derailed, compared with 7 cars derailed in a bearing-failure-caused derailment. Although a bearing failure has the potential to cause a severe train derailment, 50% of them caused a single-car derailment. Because broken rails likely to pose greater risk than other causes due to its high frequency and severity, this paper focuses on modeling train derailment severity for this cause. However, the methodology can be adapted to other accident causes.

4. Data

4.1. Data source

Train derailment data were from the Rail Equipment Accident (REA) database maintained by the Federal Railroad Administration (FRA) of U.S. Department of Transportation (U.S. DOT). Railroads in the U.S. are required to submit detailed accident reports on all accidents that exceeded a specified monetary threshold of damage costs to on-track equipment, signals, track, track structures and roadbed. The reporting threshold is periodically adjusted, and has increased from \$5700 in 1990 to \$9400 in 2011 (FRA, 2011). The REA database contains detailed train accident information such as total damage costs, number of cars derailed, track type, train length, derailment speed and others. In some previous studies, monetary damage has been used to assess the severity of train derailments. However, the financial cost of a derailment is subject to many variables, such as the cost difference between locomotives and railcars, or the difference in repairing regular track versus special trackwork. Instead, number of cars derailed may better represent train derailment severity under certain circumstances (Barkan et al., 2003).

4.2. Explanatory variables

Several factors may affect train derailment severity, including residual train length, derailment speed, train power distribution and proportion of loaded railcars in the train. The explanation to each variable is presented below.

4.2.1. Residual train length

Residual train length is defined as the number of railcars following the point-of-derailment (POD), where POD is the position of the first car derailed. Residual train length describes the maximum number of cars potentially subject to derailment (Saccomanno and El-Hage, 1989, 1991). Previous studies have found that a greater residual train length is expected to result in more cars derailed, given all else being equal (e.g., Saccomanno and El-Hage, 1989, 1991; Anderson, 2005; Bagheri, 2009; Bagheri et al., 2011).

4.2.2. Derailment speed

Nayak et al. (1983), Treichel and Barkan (1993), Saccomanno and El-Hage (1991), Saccomanno and El-Hage (1989), Anderson (2005), Bagheri (2009), Bagheri et al. (2011) and Liu et al. (2011) all showed a positive correlation between the mean number of cars derailed and derailment speed.

4.2.3. Distribution of train power

No previous study analyzed whether distributed train power could affect train derailment severity. In this study, freight-trains are classified by two types: (1) non-distributed-power trains with only head locomotives and (2) distributed-power trains with head-end locomotives and additional locomotives in other positions (typically in the middle and/or in the rear). A binary variable (1 represents a distributed-power train, 0 otherwise) is created to examine the hypothesis that the two types of trains do not have statistically different derailment severities.

4.2.4. Proportion of loaded cars

This is another new factor considered in this paper. The proportion of loaded cars in the train is defined as the ratio of number of loaded cars normalized by total number of cars (both empty and loaded) in the train. The null hypothesis is that a train carrying a larger proportion of loaded cars may derail more cars. A larger proportion of loaded cars in the train may also indicate greater kinetic energy in the derailment, thereby causing more cars to derail, given all else being equal.

Table 1 presents some descriptive statistics of the studied explanatory variables. The Spearman correlation coefficients[5] are presented in Table 2. The significant correlation is between train power distribution and the proportion of loaded cars in the train, at 5% significance level. It indicates that a derailed train having a higher proportion of loaded cars is more likely to be equipped with distributed power.

5. Zero-truncated negative binomial (ZTNB) model

5.1. Model development

The number of cars derailed represents non-negative count data, whose mean value can be estimated using regression techniques. Poisson regression and negative binomial (NB) regression are among the most popular count data regression methods used in accident analysis (e.g., Maccullagh and Nelder, 1989; Miaou, 1994; Hauer, 2001; Wood, 2002, 2005; Lord et al., 2005; Lord and Mannering, 2010). The Poisson model is suitable for data whose mean is equal to its variance, whereas the NB model assumes that the Poisson mean follows a gamma distribution. The NB model has been used for analyzing over-dispersed data (the variance is

---

[4] A group of the largest railroads accounting for 97% of traffic (ton-miles) and 94% of total freight rail revenue in the U.S. (AAR, 2011).

[5] Pearson correlation coefficients were also computed and yielded similar results.