International Journal of Reliability and Applications
Vol. 12, No. 2, pp. 95-102, 2011

# The auto regression model of bus fleet failure number

**Y. Zhou***
*School of Management and Economics of University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China*

**Abstract.** This paper uses the auto regression model to modeling failure number of a bus fleet. The fitted model can be used to predict the failure number in the future. A numerical example is presented to illustrate the modeling process and the appropriateness of the fitted model. At last, some possible applications of the model are discussed.

**Key Words:** *Vehicle fleet management, auto regression model, failure number*

## 1. INTRODUCTION

Fleet management involves operation and maintenance scheduling, spare parts and inventory management, retirement decision-making and many other issues. Scientific and effective management are benefit for improving vehicle utilization and reducing operating cost.

The fleet operators typically use fleet management system to record information about fleet reliability, economy and so on. In fact, the scientific analysis of this information can provide decision support for fleet management, which will be illustrated in this paper.

For a specific vehicle, fleet management system records failure number periodically (e.g., weekly or monthly). So we could get a sample of the random failure point process, which can be denoted as $\{n_i(t_{i-1}, t_i), i = 1, 2, 3, ...\}$. Here, $t_0 = 0$ is the beginning time of the vehicle to operate, $t_i$ denote the observation time, and $n_i$ denote the failure number of the interval $(t_{i-1}, t_i)$. If $\Delta t$ ( $\Delta t = t_i - t_{i-1}$ ) is a constant, and which can be taken as a unit of time. Obviously, $n_i$ is the empirical failure rate (Coetzee, 1997) (or failure intensity), which is defined as the expected number failure per unit time.

Let $N(t)$ denote the cumulative failure number at $(0, t)$. The expectancy of $N(t)$, $E[N(t)]$, fitted by the following power-law model (Coetzee, 1997) :

* Corresponding Author.
E-mail address: *zhouwentao951@163.com*

$$E\left[N(t)\right] = \left(\frac{t+\tau}{a}\right)^{b}. \tag{1.1}$$

The intensity function is given by

$$\lambda(t) = \frac{b}{a}\left(\frac{t+\tau}{a}\right)^{b-1}. \tag{1.2}$$

Here, $\tau$ , a and b are model parameters. They can be understood it as the initial age, time between failure and the failure trend respectively. If $b>1$, the item degrades with time; if $b=1$, the failure arrivals follow a homogeneous Poisson process, whose underlying failure times are exponentially distributed; and if $b<1$ , the item improves with time. As an ageing system, the bus performance degrades with the increasing time. Hence, it appears necessary to implement preventive maintenance at an appropriate time. Maintenance optimization model (Jiang and Murthy, 2008) is used to determine the appropriate time based on a stochastic failure point process model. Failure point process model also can be used to predict failure number in the future. Literature (Jiang et al., 2010; Jiang and Zhou, 2010) used the power-law model to fit the failure number of a bus fleet, further to evaluate bus health and determine the overhaul time. Fleet management system records the failure data periodically, which forms a time-series data. Using time-series model (e.g. auto-regression model, AR for short) to modeling the data seems natural. In the literature, many scholars have used the time series model into reliability modeling and analysis (Huang et. al., 2010; Fengchun et. al., 2011; Zhongmin and Yeqing, 2010; Roe and Mba, 2009; Christian, 2008). To best of our knowledge, the AR model is scarcely used to modeling and analysis the failure number of a bus fleet. So, we try to use the AR model to model the failure number of a bus fleet. More important, the appropriateness of the two models will be distinguished.

The paper is organized as follows. Section 2 introduces the AR model. Section 3 uses the AR model to modeling the failure number of a bus fleet, and compare with the result of the power-law model. A discussion of the model applications and a brief summary will be given in Section 4.

## 2. BASIC OF THE AR MODEL

### 2.1 Time-series of failure number

Assuming a bus fleet consists of $m$ buses (with the same model), which began operating in the same route. The fleet management system records all kinds of operational information, such as mileage, fuel consumption, failure occurrence time, corrective actions, and so on. We use a vector $\left(z^{(j)}(t), j=1,2,...,k\right)$ to denote them. When we only focus on one of them, it can simply be denoted as $z(t)$ .When the time of data collection is $t_i = i\Delta t$ ( $\Delta t$ is constant), let $\Delta t = 1$ ( $\Delta t$ is a time unit), so $t_i = i$ . Therefore, the time-series data $\{z(t_i), i = 1, 2, ...\}$ can be denoted as $\{z_i, i = 1, 2, ...\}$ .In this paper, the time unit is week. $z_i$ denotes the cumulative failure number at $(0, t_i)$, so $z_i = N(t_i)$ .

## 2.2 AR model

Assuming we have had a time-series data $\{z_1, z_2, ..., z_n\}$ . With regard to the AR model, the predictive value of the next time is the liner combination of the few former true values. It is given by Hamilton (1994).

$$z_k = \beta_0 + \sum_{j=1}^{q} \beta_j z_{k-q-1+j} + \varepsilon_k, k = q+1, q+2, \dots . \tag{2.1}$$

Here, model order $q$ is a known positive integer. $\beta_0$ and $\beta_j$ are auto regression coefficients, which is independent from *n*. $\varepsilon_k$ is a normal random variable, zero mean, variance is $\sigma^2$ . Equation (2.1) can be denoted by $AR(q)$.

In order to determine the auto regression coefficients, let $k = q+1, q+2, ..., n$ successively, so we can obtain the linear regression equation about $\{\beta_j, 0 \le j \le q\}$ :

$$\begin{aligned} z_{q+1} &= \beta_0 + \sum_{j=1}^{q} \beta_j z_j + \varepsilon_{q+1} \\ z_{q+2} &= \beta_0 + \sum_{j=1}^{q} \beta_j z_{j+1} + \varepsilon_{q+2} \ . \\ &\dots\dots \\ z_n &= \beta_0 + \sum_{j=1}^{q} \beta_j z_{n-q+j-1} + \varepsilon_n \end{aligned} \tag{2.2}$$

Form Equation (2.2), $\{\beta_j, 0 \le j \le q\}$ can be obtained through the regression analysis (Microsoft Excel has the regression analysis module), and denoted as $\{b_0, b_1, b_2, ..., b_q\}$ . The accuracy of model can be measured by $\{\varepsilon_k, k = q+1, q+2, ..., n\}$ :

$$\varepsilon_k = b_0 + \sum_{j=1}^{q} b_j z_{k-q+j-1} - z_k, k = q+1, q+2, ..., n . \tag{2.3}$$

According to Equation (2.3), $\sigma^2$ can be estimated by

$$\sigma^2 = \frac{1}{n-q} \sum_{k=q+1}^{n} \varepsilon_k^2 . \tag{2.4}$$

The smaller $\sigma$ means the more accuracy.

## 2.3 The determine of $q$

Generally speaking, with the increase of $q$, $\sigma$ will be smaller, but the model is more complex. Therefore, determining the appropriate $q$ is the key of the modeling processes. In this paper, we determine $q$ based on $P$ value of the regression analysis. A smaller $P$ value means the linear relation between variables more significant. The threshold of $P$ value is usually given as 0.05. According to *AR*(*q*), the number of *P* value is *q+1*, and let $p_m$ denote the biggest *P* value. With the increase of *q*, more variables will be introduced, and lead to $p_m > 0.05$ . Until $q_0$ can satisfy the following relation:

$$p_m(q) < 0.05, q \le q_0; p_m(q_0+1) > 0.05 . \tag{2.5}$$

So the $q_0$ is the true value which we are looking for.

### 2.4 Prediction

Once the model parameters were identified, the next predictive value is:

$$z_{n+1} = b_0 + \sum_{j=1}^{q_0} b_j z_{n-q_0+j} . \tag{2.6}$$

When we predict the values of $t = n+2, n+3, \ldots$, the predictive values of $t = n+1, n+2, \ldots$ can be taken as observed value. When there are more observational data, the model can be updated to make better prediction accuracy.

## 3. AR MODEL OF BUS FLEET FAILURE NUMBER

### 3.1 Background and data

A certain fleet of 22 buses (with the same model) began operating in one of the routes on August 24th, 2005. Upon an operational failure, the bus is restored by a corrective repair, which can be deemed a minimal repair though certain opportunistic maintenance actions may be combined. We have collected the buses operational data (failure number per week and maintenance cost of each bus and fuel consumption of the whole fleet) in duration of 175 weeks (September 1st, 2006, when the management system began operating, to December 31st, 2009). In this paper, we focus on the analysis of failure number. $z_i$ denote the failure number at interval $(0,i)$, $n_i$ denote the failure number at interval $(i-1,i)$:

$$n_i = z_i - z_{i-1} . \tag{3.1}$$

The failure observations are displayed in Table 3.1.

In order to test the accuracy of the fitted AR model, we use the previous 160 data to modeling the AR model, and predict the following 15 week's failure number, then compared with the actual observations. The comparison model is the power-law model.

### 3.2 Modeling process

According to the abovementioned method, we can estimate the model parameters using the regression module in Microsoft Excel. Let *q*=1, 2 and 3, the model parameters and *P* values are given in Table 3.2. According to Table 3.2, we have $q_0 = 2$ . We also fit the data using the power-law model. The model parameters are given in the last row of

Table 3.2. The fitting results of the two models are shown in Figure 3.1. The figure shows that the power-law model describes the average failure behavior of the failure process, and the AR model reflects the dynamic behavior of the failure process.

**Table 3.1.** Failure number of the bus fleet

| $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ | $i$ | $z_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 4 | **21** | 182 | **41** | 384 | **61** | 672 | **81** | 984 | **101** | 1373 | **121** | 1822 | **141** | 2284 | **161** | 2884 |
| **2** | 14 | 22 | 193 | **42** | 400 | **62** | 680 | **82** | 1001 | **102** | 1393 | **122** | 1846 | **142** | 2316 | **162** | 2916 |
| **3** | 22 | **23** | 205 | **43** | 414 | **63** | 685 | **83** | 1011 | **103** | 1420 | **123** | 1882 | **143** | 2336 | **163** | 2946 |
| **4** | 29 | **24** | 218 | **44** | 430 | **64** | 695 | **84** | 1020 | **104** | 1441 | **124** | 1904 | **144** | 2365 | **164** | 2976 |
| **5** | 43 | **25** | 225 | **45** | 441 | **65** | 712 | **85** | 1030 | **105** | 1463 | **125** | 1923 | **145** | 2404 | **165** | 3007 |
| **6** | 52 | **26** | 232 | **46** | 457 | **66** | 737 | **86** | 1050 | **106** | 1479 | **126** | 1937 | **146** | 2445 | **166** | 3031 |
| **7** | 57 | **27** | 239 | **47** | 475 | **67** | 754 | **87** | 1070 | **107** | 1504 | **127** | 1955 | **147** | 2490 | **167** | 3055 |
| **8** | 67 | **28** | 245 | **48** | 492 | **68** | 763 | **88** | 1088 | **108** | 1522 | **128** | 1976 | **148** | 2511 | **168** | 3090 |
| **9** | 77 | **29** | 253 | **49** | 515 | **69** | 780 | **89** | 1112 | **109** | 1547 | **129** | 2000 | **149** | 2538 | **169** | 3122 |
| **10** | 85 | **30** | 265 | **50** | 530 | **70** | 800 | **90** | 1143 | **110** | 1563 | **130** | 2029 | **150** | 2561 | **170** | 3144 |
| **11** | 93 | **31** | 276 | **51** | 549 | **71** | 815 | **91** | 1158 | **111** | 1580 | **131** | 2051 | **151** | 2606 | **171** | 3170 |
| **12** | 102 | **32** | 287 | **52** | 558 | **72** | 832 | **92** | 1178 | **112** | 1590 | **132** | 2080 | **152** | 2650 | **172** | 3204 |
| **13** | 106 | **33** | 295 | **53** | 572 | **73** | 848 | **93** | 1202 | **113** | 1619 | **133** | 2101 | **153** | 2683 | **173** | 3231 |
| **14** | 110 | **34** | 308 | **54** | 581 | **74** | 865 | **94** | 1219 | **114** | 1645 | **134** | 2125 | **154** | 2721 | **174** | 3258 |
| **15** | 120 | **35** | 322 | **55** | 595 | **75** | 882 | **95** | 1250 | **115** | 1676 | **135** | 2145 | **155** | 2735 | **175** | 3293 |
| **16** | 126 | **36** | 330 | **56** | 615 | **76** | 895 | **96** | 1269 | **116** | 1697 | **136** | 2163 | **156** | 2760 | | |
| **17** | 136 | **37** | 342 | **57** | 625 | **77** | 910 | **97** | 1291 | **117** | 1718 | **137** | 2194 | **157** | 2785 | | |
| **18** | 145 | **38** | 350 | **58** | 634 | **78** | 934 | **98** | 1319 | **118** | 1749 | **138** | 2217 | **158** | 2817 | | |
| **19** | 153 | **39** | 360 | **59** | 648 | **79** | 950 | **99** | 1334 | **119** | 1773 | **139** | 2239 | **159** | 2835 | | |
| **20** | 170 | **40** | 367 | **60** | 663 | **80** | 969 | **100** | 1354 | **120** | 1799 | **140** | 2259 | **160** | 2859 | | |

**Table 3.2.** Estimated $AR(q)$ model parameters and $p_m$ values

| $q$ | $b_0$ or $a$ | $b_1$ or $b$ | $b_2$ or $\tau$ | $b_3$ | $\sigma$ | $p_m$ |
|---|---|---|---|---|---|---|
| 1 | 9.0693 | 1.0079 | | | 5.5443 | 0.0000 |
| 2 | 7.1080 | -0.2196 | 1.2258 | | 5.4295 | 0.0065 |
| 3 | 6.6896 | -0.0635 | -0.1446 | 1.2137 | 5.4354 | 0.4470 |
| Model(1) | 2.1414 | 1.8006 | 17.3879 | | 17.4381 | |

### 3.3 Comparison of prediction results

Now, we apply the two fitted model to predict the failure number of 161-175 weeks. The predictive value, observed values and their relative errors are displayed in Figure 3.1 and Figure 3.2. According to Figure 3.1 and Figure 3.2, we found that: When prediction interval is less than 5, the AR model prediction errors are less than the power-law model. And when the prediction interval is greater than 5, the power-law model prediction is more accuracy. In other words, the AR model is applicable to short-term prediction (for example, less than one month); and the power-law model is applicable to long-term prediction.

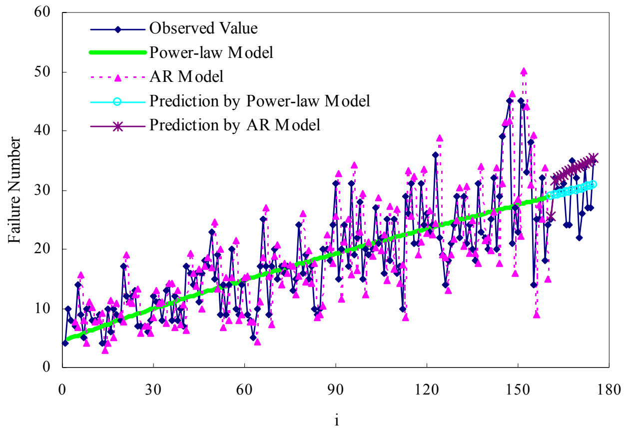


**Figure 3.1.** The fitting results based on the data of 160 weeks ago

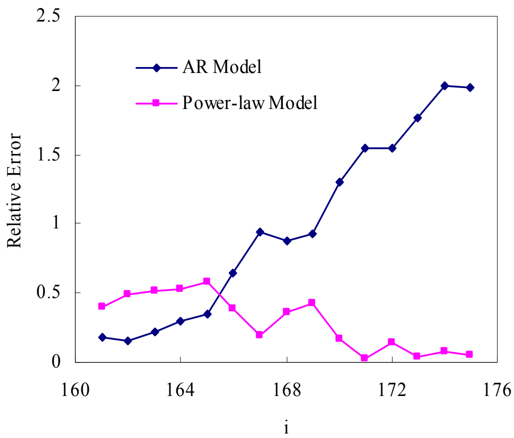


**Figure 3.2.** The relative errors of the fitted models

### 3.4 Model update

If we update the model every few weeks, we can obtain more accuracy results. Where the longer cycle is, the greater relative errors are. The Table 3.3 proofs it. Thus, according to the allowed error, we can determine an appropriate update cycle.

**Table 3.3.** The fitted model parameters based on the data of 165 and 170 weeks ago

| n | $b_0$ | $b_1$ | $b_2$ | $P_m$ |
|---|---|---|---|---|
| 160 | 7.1080 | -0.2196 | 1.2258 | 0.0065 |
| 165 | 7.1392 | -0.2260 | 1.2320 | 0.0041 |
| 170 | 7.3234 | -0.2323 | 1.2379 | 0.0030 |
| 175 | 7.3717 | -0.2371 | 1.2425 | 0.0019 |
| Relative error,%, n=165 | 0.4389 | 2.9144 | 0.5058 | |
| Relative error,%, n=170 | 2.5801 | 2.7876 | 0.4789 | |
| Relative error,%, n=175 | 0.6595 | 2.0663 | 0.3716 | |

## 4. APPLICATION AND CONCLUSIONS

In the above mentioned numerical example, the AR model is applicable to short-term prediction. The prediction of the failure number also has the following applications:

(1) Planning and scheduling the operation and production;
(2) Preparation of maintenance, such as pre-order spare parts;
(3) Capital budget;
(4) Determine the occasion of preventive maintenance, and so on.

For example, supposing the average maintenance cost is 300 RMB, and the downtime is 1 hour. In the following 4 weeks after the 175-th weeks, there will be 101 failures. In that case, the maintenance cost and downtime will be 30300 RMB, and 101 hours respectively. About the application of maintenance decision-making, the reader may refer to Jiang et al., (2010) and Jiang and Zhou (2010).

In short, we have used the AR model to modeling the failure number of a certain bus fleet, and compared with the power-law model. The main findings are:

(1) The power-law model describes the average failure behavior of the failure process, and the AR model reflects the dynamic behavior of the failure process;
(2) The AR model is applicable to short-term prediction; and the power-law model is applicable to long-term prediction;
(3) According to the allowable error level, we can determine the model update cycle.

## REFERENCES

Coetzee, J. L. (1997). The role of NHPP models in the practical analysis of maintenance failure data, *Reliability Engineering and System Safety*, **56**, 161-168.

Jiang, R. and Murthy, D.N.P. (2008). *Maintenance: Decision Models for Management*, Science press, Beijing.

Jiang, R., Mi, M. and Zhou, Y. (2010). Analysis of bus operational performance and overhaul decision, *The IEEE Conference on Engineering and Business Management*, Chengdu, 5734-5737.

Jiang, R. and Zhou, Y. (2010). Failure-counting based health evaluation of a bus fleet. *2010 Prognostics and System Health Management Conference*, MU3013, 1-4.

Tingting Huang, Li Wang and Tongmin Jiang (2010). Prognostics of products using time series analysis based on degradation data, 2010 *Prognostics and System Health Management Conference*, MU3106, 1-5.

Lin Fengchun, Chen Yunxia and Kang Rui (2011). Degradation analysis method based on regretssion time series model under equal and unequal variances, *2011 Prognostics and System Health Management Conference*, MU3106, 1-7.

Chen Zhongmin and Wu Yeqing (2010). The application of theory and method of time series in the modeling of software reliability, *2010 Second International Conference on Information Technology and Computer Science*, 340-343.

Roe, S. and Mba, D. (2009). The environment, international standards, asset health management and condition monitoring: an integrated strategy, *Reliability Engineering and System Safety*, **94**, 474-478.

Christian H. Weib (2008). The combined INAR(p) models for time series of counts, *Statistics and Probability Letters*, **78**, 1817-1822.

Hamilton, J. D. (1994). *Time series analysis*, Princeton university press, New Jersey.