1994

An Investigation of the Statistical Relationship between the 345 kv Transmission Line Length and the Outage Rate

Sohn, S.Y.

World Scientific

Sohn, S.Y., "An Investigation of the Statistical Relationship between the 345 kv Transmission Line Length and the Outage Rate," International Journal of Reliability, Quality and Safety Engineering, Vol. 1, pp. 85-94, 1994.
http://hdl.handle.net/10945/60812

International Journal of Reliability, Quality and Safety Engineering
Vol. 1, No. 1 (1994) 85–94

AN INVESTIGATION OF THE STATISTICAL RELATIONSHIP BETWEEN THE 345 KV TRANSMISSION LINE LENGTH AND THE OUTAGE RATE

SO YOUNG SOHN
Department of Operations Research
Naval Postgraduate School, Monterey, CA 93943, U.S.A.

Received 26 July 1993
Revised 28 December 1993

Finding the factors that are related to varying transmission outage rates has been one of the major concerns of the reliability practitioners in the electric power industry. One of the potentially influential factors is the length of the transmission line. In this paper, a random effects Poisson regression model is used to quantify the relationship between the outage rate to the transmission line length. The method suggested is applied to analyze two sets of the 345 KV transmission line outage data of the Commonwealth Edison Company: group 1 contains transmission lines that had been installed before January 1974; group 2 consists of the lines that have been installed since January 1974. Results indicate that in both cases there is a significant log linear relationship between the transmission line outage and the line length. However, the annual outage rate elasticity of transmission lines of group 2 turns to be 0.33 while that of group 1 is 0.64. This would imply apparent quality improvement on the transmission lines in group 2.

Keywords: Transmission Line Outage Rates; A Random Effects Poisson Regression; Lognormal Distribution.

Introduction

Finding the relationship between the outage rates of transmission lines and the associated line characteristics is one of the major concerns of reliability practitioners in the electric power industry. The mathematical formulation of such a relation can be described using the following approach.

Consider a collection of N transmission lines that independently experiences n_i outages in accordance with Poisson processes during time t_i, each with outage rate parameter λ_i for line i, $i = 1, \ldots, N$. However, values of λ_i would vary randomly. Part of the variation could be explained by the associated line characteristics, x_i, such as transmission line length, quarry exposure and the terminal configuration. What follows is a regression model for λ_i against those covariates, x_i. To find the relationship between λ_i and x_i, the observed values of λ_i and x_i are necessary. Since λ_i is unobservable and random, it is often replaced with $\hat{\lambda}_i = n_i/t_i$ with the to be computationally burdensome and difficult to implement. In an effort to improve on these solutions, we consider a two-stage estimation where within-individual parameters are separately estimated from the between-individual regression model.

A Two-Stage Estimation Method

First, in the within-individual model, the annual rate of outages is estimated as $\hat{\lambda}_i = n_i/t_i$. The $\hat{\lambda}_i$ replaces the unobservable λ_i in the between-individual model (3). As a result of this replacement, the estimation error δ_i, that is independent of ϵ_i, is added in the model:

$$\log(\hat{\lambda}_i) = \beta_1 + \beta_2 \log x_{i2} + \cdots + \beta_p \log x_{ip} + \epsilon_i + \delta_i \,. \tag{6}$$

Using the delta method, $\log(\hat{\lambda}_i)$ can be regarded as having approximate normal distribution with mean $\beta_1 + \beta_2 \log x_{i2} + \cdots + \beta_p \log x_{ip}$ and variance $(\sigma^2 + 1/n_i)$ based on the following derivation:

$$E(\log \hat{\lambda}_i | \lambda_i) \simeq E(\log \lambda_i + (\lambda_i - \hat{\lambda}_i)/\lambda_i) = \log(\lambda_i); \text{ and}$$

$$V(\log \hat{\lambda}_i | \lambda_i) = E(\log \hat{\lambda}_i - \log \lambda_i | \lambda_i)^2 \simeq E((\lambda_i - \hat{\lambda}_i)/\lambda_i)^2 = 1/n_i \,.$$

Since the variance is not constant due to a different frequency of failure n_i, the weighted least square (WLS) method is then used to estimate β using $\hat{\beta}(\sigma^2) = (Z'WZ)^{-1}(Z'W(\log \hat{\lambda}))$ and $\mathrm{cov}(\hat{\beta}(\sigma^2)) = (Z'WZ)^{-1}$ where Z is an $N \times p$ covariate matrix of $z_{ij} = \log x_{ij}$; W is an $N \times N$ diagonal matrix consisting of diagonal elements $w_i = 1/(\sigma^2 + 1/n_i)$; and $\log(\hat{\lambda})$ is an $N \times 1$ column vector of $\log(\hat{\lambda}_i)$. Often σ^2 in W is not known and has to be estimated using the available data. The ML estimates, $\hat{\sigma}^2$ can be obtained by finding the σ^2 which maximizes the following log likelihood function $L(\sigma^2, \beta'; \hat{\lambda}, x, n)$:

$$-0.5 \sum_{i=1}^{N} \log(\sigma^2 + 1/n_i) - 0.5 \sum_{i=1}^{N} (\log \hat{\lambda}_i - z_i \beta)^2/(\sigma^2 + 1/n_i) \tag{7}$$

where z_i is a $1 \times p$ vector of $(1, \log x_{i2}, \ldots, \log x_{ip})$. Since there is no closed form solution for $\hat{\sigma}^2$, an iterative approach can be employed. For the initial values of parameters, ordinary least square estimates, $\hat{\beta}^0 = (Z'Z)^{-1}(Z'(\log \hat{\lambda}))$ and $\hat{\sigma}_0^2 = \sum_{i=1}^{N} (\log \hat{\lambda}_i - z_i \hat{\beta}^0)^2/(N-p)$ are used. Once the ML estimate, $\hat{\sigma}^2$ is obtained, the ML estimate $\hat{\beta}(\hat{\sigma}^2) = (Z'\hat{W}Z)^{-1}(Z'\hat{W}(\log \hat{\lambda}))$ where $\hat{W}$ consists of the diagonal elements $\hat{w}_i = 1/(\hat{\sigma}^2 + 1/n_i)$. The resulting $\hat{\lambda}_{i'}^{R} = \exp(z_{i'} \hat{\beta}(\hat{\sigma}^2))$ can be used for the prediction of $\lambda_{i'}$ for the new line i'. An interval estimate can also be obtained using $\hat{\mathrm{cov}}(\hat{\beta}(\hat{\sigma}^2)) = (Z'\hat{W}Z)^{-1}$.

A Case Study

A two-stage Poisson regression model is used to analyze two sets of the outage data obtained from the 345 KV transmission lines of the Commonwealth Edison Company. Group 1 covers 112 transmission lines that were active any time between 1965 and 1976, which is identical to a set of lines presented in Table A.1 of Alsammarae's study.[5] Additionally, we analyze group 2 that consists of lines which have started their service since January 1974 and were in operation any time between 1982 to 1991. From both group 1 and 2, we eliminate lines whose operation period does not exceed 3 years. This is to increase the accuracy of the estimation. The two groups can, in turn, be distinguished in terms of not only their operation period [(1965–1976) for group 1 and (1982–1991) for group 2] but also their starting time of service [(before January 1974) for group 1 (since January 1974) for group 2]. We carry out a two-stage analysis using the remaining transmission lines (61 in group 1 and 74 in group 2).

First, in the within-individual model, each individual outage rate is estimated using $\hat{\lambda}_i = n_i/t_i$. Tables 1 and 2 summarize the outage history of the transmission lines of groups 1 and 2, respectively, along with the line length (x_{i2}).

Table 1. Summary of transmission outage data: 61 lines in group 1 observed during 1965–1976.

	Unit	Mean	Std Dev	Max	Min
n_i		11.10	12.28	63.00	0.00
t_i	year	5.87	2.04	12.00	3.26
$\hat{\lambda}_i$	per year	1.86	1.83	8.25	0.00
x_{i2}	mile	32.06	26.39	99.5	4.60

Table 2. Summary of transmission outage data: 74 lines in group 2 observed during 1982–1991.

	Unit	Mean	Std Dev	Max	Min
n_i		11.50	12.07	73.00	0.00
t_i	year	8.95	2.08	10.00	3.02
$\hat{\lambda}_i$	per year	1.28	1.29	8.04	0.00
x_{i2}	mile	25.99	23.61	126.30	3.50

The relative frequency of $\hat{\lambda}_i$ in group 1 is illustrated in Fig. 1. Apparent outliers ($\hat{\lambda}_i > 8$) support the selected use of the lognormal super-population. For the preliminary analysis of the between-individual model, we plot $\log(\hat{\lambda}_i)$ against $\log x_{i2}$ for group 1 in Fig. 2.

In the process of obtaining estimates, $\hat{\beta}(\hat{\sigma}^2)$ and $\hat{\sigma}^2$, we replace $\log\hat{\lambda}_i$ with $\log(1/3t_i)$ when $\hat{\lambda}_i = 0$ following the method used in Gaver *et al.*[8] The maximum likelihood estimates, $\hat{\beta}(\hat{\sigma}^2)$ and $\hat{\sigma}^2$, are obtained using a standard computer optimization package GAMS.[9] The fitted model is as follows:

$$\log\hat{\lambda}_i^R = -1.6288 + 0.6379\log x_{i2}\,. \qquad (8)$$

Parameter estimates in Eq. (8) are significant at 1% level and the fitted $\log\hat{\lambda}_i^R$ is overlaid to the $\log\hat{\lambda}_i$ against $\log x_{i2}$ as a solid line in Fig. 2. For a diagnostic check,

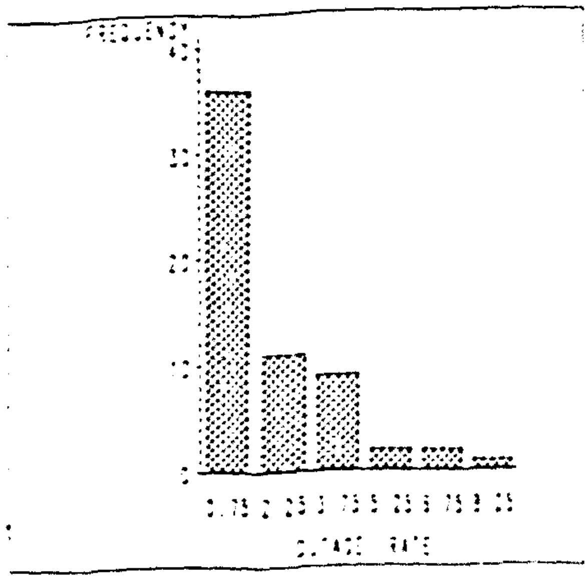

Fig. 1. Relative frequency of the annual outage rate per mile, $\hat{\lambda}$: group 1.

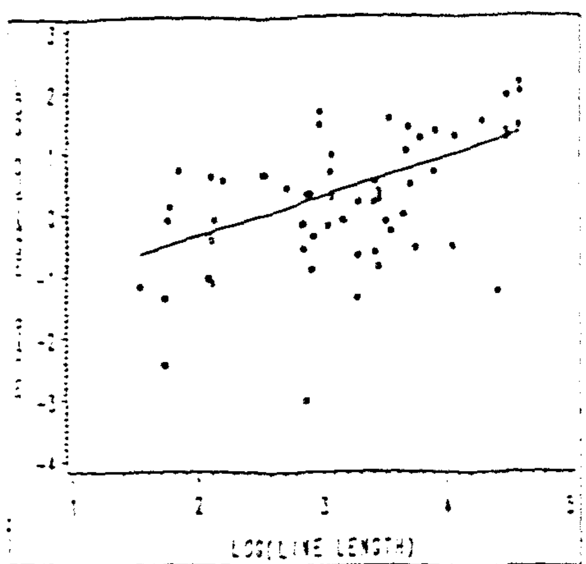

Fig. 2. The relationship between the estimated outage rate, $\log \hat{\lambda}$ and the line length, $\log x$: group 1.

the weighted residuals $\sqrt{w_i}\,(\log \hat{\lambda}_i + 1.6288 - 0.6379 \log x_{i2})$ are analyzed using the PROC UNIVARIATE of the statistical package, SAS.[10] A normality test is passed for this example based on Kolomogorov D statistic at the 5% significance level.

A plot of $\log \hat{\lambda}_i$ against $\log x_i$, given in Fig. 2, may not appear to strongly back up a significant log-linear relationship between the outage rate and the line length. Therefore we add Fig. 3 in which $\sqrt{\hat{w}_i}\,\log \hat{\lambda}_i$ is plotted against $\sqrt{\hat{w}_i}\,\log x_i$. Notice that when each observation i might be associated with different variability $(\sigma^2 + 1/n_i)$, fitting model (6) is the same as fitting

$$\sqrt{\hat{w}_i}\,\log \hat{\lambda}_i = \beta_1 \sqrt{\hat{w}_i} + \beta_2 \sqrt{\hat{w}_i}\,\log x_{i2}, \ldots, + \beta_p \sqrt{\hat{w}_i}\,\log x_{ip} + \xi_i \tag{9}$$

where $\xi_i \sim N(0,1)$. Figure 3 gives a better insight concerning the relationship between $\log \hat{\lambda}_i$ and $\log x_i$ than Fig. 2 which does not reflect different weights associated with each observation.

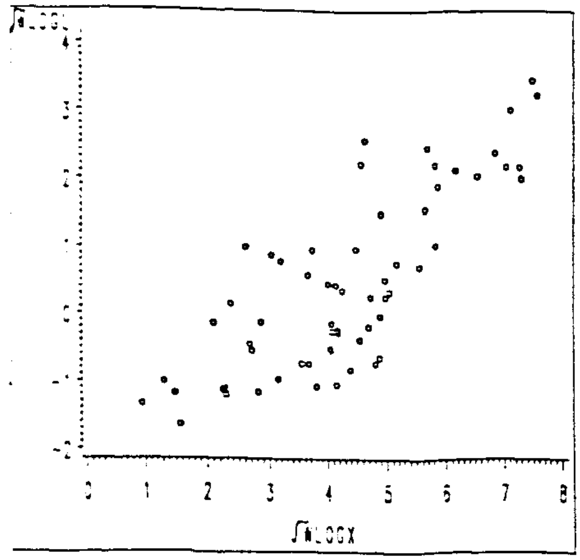

Fig. 3. The relationship between the $\sqrt{\hat{w}} \log \hat{\lambda}$ and $\sqrt{\hat{w}} \log x$: group 1.

The useful measure of the responsiveness of the annual outage rate of a particular transmission line to its line length is the elasticity defined as the ratio of the relative change in the annual outage rate to the relative change in the length of the line $((\Delta \hat{\lambda}_i/\hat{\lambda}_i)/(\Delta x_i/x_i))$. Fitted model (8) implies the following relationship between $\hat{\lambda}_i$ and the line length x_i:

$$\hat{\lambda}_i^R = \exp(-1.6288) x_i^{0.6379} . \tag{10}$$

Evaluation of $(\Delta \hat{\lambda}_i/\hat{\lambda}_i)/(\lambda x_i/x_i)$ based on model (8) brings estimated annual outage rate elasticity of transmission line length in group 1 to about 0.64.

Next, it is observed that model (8) performs better than the fitted model in Alsammarae's study[5] in terms of the mean squared error (MSE). The MSE of model (8) is 2.0069 while that of $\hat{\lambda}_i^R = 1.4813 + 0.0153 x_i$ in his study[5] is 3.5224.

In addition to the line length, we analyze some other covariates (quarry exposure, configuration of terminal and three terminal stations) which were used by Schneider *et al.*[4] However, they do not show a significant contribution to fit the log outage rates.

We apply a two-stage estimation procedure to group 2. Corresponding plots to Figs. 1, 2 and 3 for group 1 are given in Figs. 4, 5 and 6, respectively. The fitted model is as follows:

$$\log \hat{\lambda}_i^R = -0.9286 + 0.3347 \log x_i . \tag{11}$$

The estimated parameters are significant at 2%. The estimated annual outage rate elasticity of transmission line length in group 2 is about 0.33, that is less than 0.64

for group 1. It would imply a technological improvement made on the transmission lines that were in service after 1974 compared to that of group 1. The estimated annual outage rate elasticity of transmission line length can be compared to those of other companies and it can be used as an input for performance optimization of the transmission lines.

Although model (9) provides information concerning the relationship between the outage rate and the transmission line, it is noted that the transmission line alone ($R^2 = 33\%$) cannot explain overall variation of the outage rates. Additional characteristics of the transmission lines analyzed in this study (quarry exposure, terminal configuration and three terminal stations), however,

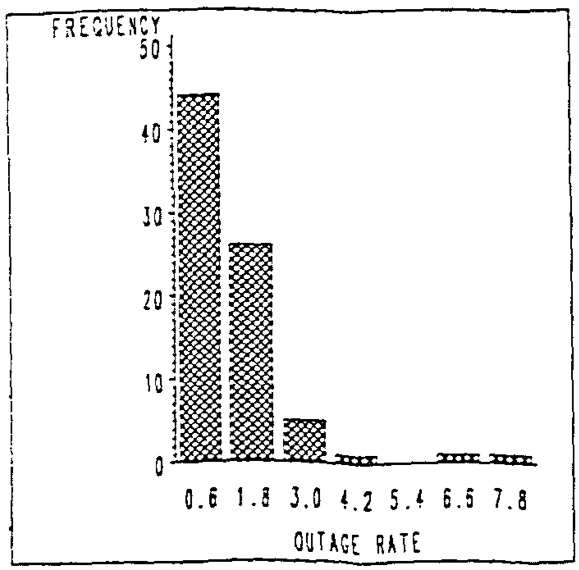

Fig. 4. Relative frequency of the annual outage rate per mile, $\hat{\lambda}$: group 2.

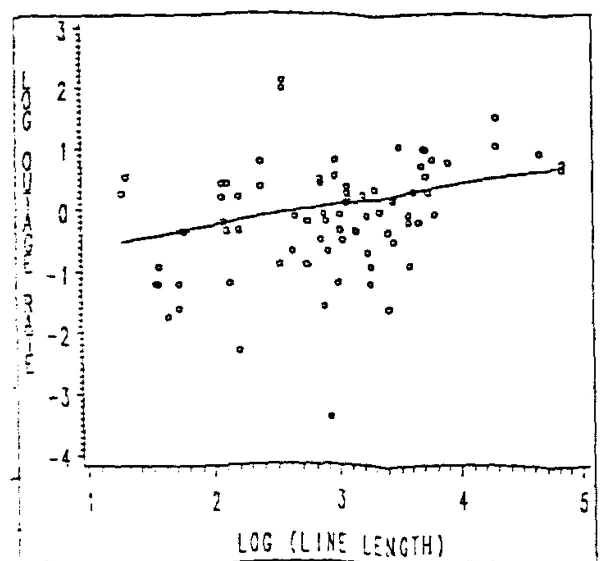

Fig. 5. The relationship between the estimated outage rate $\log \hat{\lambda}$ and the line length, $\log x$: group 2.

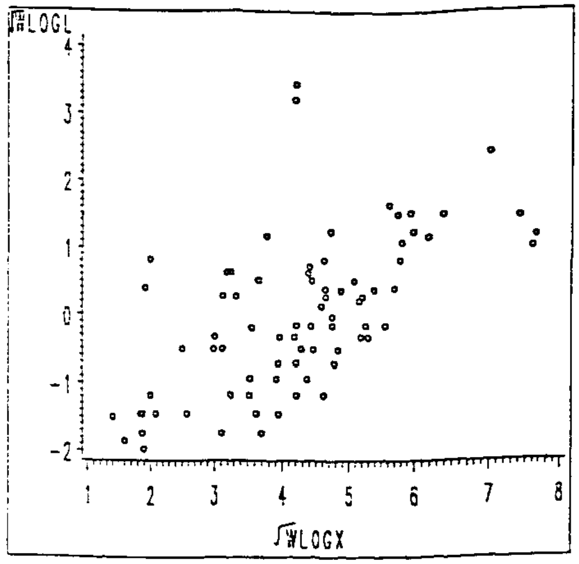

Fig. 6. The relationship between the $\sqrt{\hat{w}}\log\hat{\lambda}$ and $\sqrt{\hat{w}}\log x$: group 2.

turns out to be statistically insignificant in terms of explaining the variation in $\log\hat{\lambda}_i$. By including extra variables such as the weather index, R^2 could have been increased. In order to reflect the effect of weather conditions on the transmission outage rate, an analysis of quarterly outage rate would be a better option than that of the annual outage rate.

Conclusions

A random effects Poisson regression model is employed to relate the annual outage rate of transmission line to the line characteristics. Unlike the previous methods used for the same purpose, the random effects model suggested in this article accommodates both the random error as well as the estimation error of the outage rate. A two-stage estimation method is developed to estimate the outage rate of the transmission line in terms of the corresponding line length. A log-linear model employed to describe the relationship between the outage rate and the line length has several advantages over the conventional linear model. First, it guarantees a non-negative estimate of the outrage rate. Secondly, the estimated coefficient can be interpreted as the outage rate elasticity with respect to each covariate.

The estimation method suggested in this paper is applied to the two groups of outage data on the 345 KV transmission lines of the Commonwealth Edison Company. The results obtained from this analysis provide the electric power industry with the following insights: (1) there is a significant log-linear relationship between the annual outage rate and the line length; (2) for the transmission lines manufactured after 1974, provided that they were in use for more than 3 years, as the line length increases by 1% from the current value, the outage rate increased by an average of 0.33% from the observed value; (3) there has been a quality improvement on the transmission lines manufactured after 1974 by reducing the elasticity from

0.64 to 0.33. The reduced elasticity might be due to several changes made in these periods in terms of vendor selection, design and construction mode. These conclusions are based on the analysis of data provided from one company. In order to draw more general conclusions, the analysis of various transmission lines provided by different vendors/years is recommended. When such information is available, a similar model to (6) can be used with a set of more extensive candidate covariates such as transmission line type (e.g., fossil unit or nuclear unit), various vendors, manufacturing year, and average operating conditions, etc.

Results of such a random effects Poisson regression analysis can be used as valuable inputs to the transmission line quality control. Based on the estimated relationship between the outage rate and covariates, one can select the appropriate vendor who has supplied better quality lines than the others. At the same time one can reassess the purchasing cost and take actions for the controllable operating conditions to decrease the potential outage rate.

Concerning the aspect of the data analysis, in order to increase the accuracy of the prediction, it is recommended to analyze the quarterly outage rates of transmission lines as a response variable. The advantage of using quarterly data to annual data would be the facilitation of the possible seasonal effects on the outage rate in the model.

Acknowledgment

I would like to express my appreciation to the editor and the two referees for their insightful comments and to Professor Mainak Mazumdar for introducing me to the problem dealt within this paper. I also thank Mr. Peter Burke from the Commonwealth Edison Company for the data set he provided me with and several comments he made on the results of the analysis. This paper is prepared in conjunction with research funded by the Naval Postgraduate School.

References

1. A. D. Paton, *IEEE Transactions on Power Apparatus and Systems* **PAS-87**, 84 (1968).
2. J. A. Marshall, M. Mazumdar and D. A. McCutchen, *Component Outage Data Analysis Methods, Vol. 1: Overview and Special Applications* (EPRI EL-1980, 1981).
3. G. L. Landgren, A. W. Schneider and M. P. Bhavaraju, *Predicting Transmission Outages for System Reliability Evaluations, Vol. 1: Methodology and Validation* (EPRI EL-3880, 1985).
4. A. W. Schneider, G. L. Landgren and M. P. Bhavaraju, *IEEE Transactions on Power Apparatus and Systems PAS-104* (5), 1084 (1985).
5. A. J. Alsammarae, *IEEE Transactions on Reliability* **R-35** (4), 363 (1986).
6. J. F. Lawless, *Journal of American Statistical Association* **82** (**399**), 808 (1987).
7. E. F. Vonesh and R. L. Carter, *Biometrics* **43**, 617 (1987).
8. D. P. Gaver, P. A. Jacobs and I. G. O'Muricheartaigh, *Communications in Statistics-Theory and Methodology* **19** (**10**), 3779 (1990).
9. GAMS, *GAMS User's Guide* (Scientific Press, Redwood City, CA, 1988).
10. SAS, *SAS User's Guide*, version 5 edition (SAS Institute, Cary, NC, 1989).