Investigating the Validity of Safety Performance Function in Korean Expressways

Dongwoon KANG [a], Shin Hyoung Park [b], Dong-Kyu Kim [c], Seung-Young KHO [d], Sungmo Rhee [e]

[a] *Department of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Korea,151-744; E-mail: dudandy@snu.ac.kr*
[b] *Korea Expressway Corporation Research Institute, 50-5 Sancheok-ri, Dongtan-Myeon, Hwaseong-si, Gyeonggi-do, Korea, 445-812; E-mail: shinhpark@ex.co.kr*
[c] *Integrated Research Institute of Construction and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Korea,151-744; E-mail: kimdk95@snu.ac.kr*
[d] *Same as the first author; E-mail: sykho@snu.ac.kr*
[e] *Same as the third author; E-mail: rheesm@snu.ac.kr*

Abstract: This study aims to perform the predictability validation and the cross validation based on a variety of safety performance functions (SPF). In order to do this, relationship between the number of crashes and exposures, distributions of disturbance and various functions formulated including the empirical Bayes estimation were investigated. On the basis of the estimated SPFs, the predictability validation and the cross validation were practiced by using the mean absolute deviation (MAD). As a result from both validations, the power function with exposures fit well and the results of the empirical Bayes estmation were the most satisfying.

Keywords: Crash Safety Performance Function, Empirical Bayes, Cross Validation, Functional Form, Crash Prediction

1. INTRODUCTION

Over the past years, various methods and measures have been researched regarding safety improvements. Previous studies showed the calculation of expected safety performance by comparing the past and present safety performances of certain sites. Based on this calculation, hazardous sites were selected; this was the beginning of the safety performance function (SPF), demonstrating a relation between the number of crashes and exposure (Hauer, 1995).

Previous study revealed that the distribution of crash occurrences was not a linear regression because it was positively skewed (Jovanis and Chang, 1986). Also, Miaou and Lum demonstrated that the linear regression models couldn't describe the adequately random, discrete, nonnegative, and typically sporadic vehicle accident events (Miaou and Lum, 1993). After these studies, the Poisson regression models were employed for estimating the SPFs since it described more correctly than the linear regression model (Kraus *et al.*, 1993; Khan *et al.*, 1999). However, the Poisson regression model has a limitation in which the variance is equal to the mean (Dean and Lawless, 1989). In the observed distribution of crash occurrences, the variance usually exceeds the mean; this is called over-dispersion. The negative binomial (NB) regression model, which can explain the over-dispersion, is widely used (Lu *et al.*, 2012). The NB regression was recently used by Kim *et al*. (2013). On the other hand, network screening and the before-after evaluation are important tasks in road safety management. The empirical Bayesian, which performs both network screening and the before-after evaluation, was applied by Hauer (1997) and it has been used for the past two decades (Persaud *et al*., 2010).

The objective of this paper is to perform the predictability validation and the cross validation based on a variety of functional forms to predict the number of crashes. Relationship between the number of crashes and exposures, distributions of disturbance and various functions formulated including the empirical Bayes estimation were investigated. To compare the prediction performance, four functional forms, linear regression, Poisson regression, NB regression, and empirical Bayes estimation, were used.

Models were estimated by using the data collected on three Korean expressways (Gyeongbu, Seohaean and Honam). There are two analysis: the predictability validation and the cross validation. For the predictability validation, the SPFs are estimated using the data from 2005 to 2007. Then, the number of crashes in 2008 - 2010 is predicted by the SPFs. The mean absolute deviation (MAD) is used to compare the difference between the predicted values and observed values in 2008 - 2010. In the cross validation, half of the data from 2008 to 2010 are used to estimate the SPFs. Then, cross validation is practiced using the MAD by comparing the predicted and observed values at remaining half of data from 2008 to 2010.

2. METHODOLOGY

2.1 Safety Performance Function (SPF)

A SPF is a mathematical function that describes the relation between the number of crashes per year and the measure of exposure (Hauer, 1995). Traffic, area type, number of lanes, lane width, shoulder width, horizontal and vertical curves influence the predicted number of crashes (AASHTO, 2010). The SPFs including both roadway geometric characteristics and traffic characteristics as variables are referred to inclusive SPFs. The equation to predict the number of crashes on a road way segment is (Alluri and Ogle, 2012):

$$N_{predicted} = e^{\alpha} \times AADT^{\beta_1} \times V_1^{\beta_2} \times V_2^{\beta_3} \times V_3^{\beta_4} \times V_4^{\beta_5} \times \cdots\cdots \times V_{n-1}^{\beta_n} \tag{1}$$

where,

$N_{predicted}$: the number of predicted crashes per mile per year,

$AADT$: annual average daily traffic (veh/day),

α, β_i : regression coefficients (i = 1, …, n), and

V_i : independent variables (i = 1, …, n-1),
(i.e. roadway characteristics that influence crash frequency).

The limit of inclusive SPFs is that the variables might be correlated. To avoid this problem, the SPFs based on AADT only for each roadway segment are being practiced (Lu *et al*., 2013). The equation used on roadway segments is (Alluri and Ogle, 2012):

$$N_{predicted} = e^{\alpha} \times AADT^{\beta} \tag{2}$$

where,

$N_{predicted}$: the number of predicted crashes per mile per year,

$AADT$: annual average daily traffic (veh/day), and
α, β : regression coefficients.

2.2 Empirical Bayes Estimation

In this chapter, the equation form and notation are followed Persaud *et al*. (2010). In the empirical Bayes estimation, the difference between the expected and observed number of crashes at a location is:

$$\lambda - \pi \tag{3}$$

where,
λ : the number of expected crashes in the after period without treatment,
π : the number of observed crashes in the after period.

When the expected value is estimated, the regression-to-the-mean (RTM) is accounted for using the empirical Bayes estimation. In the empirical Bayes estimation, the expected number of crashes in each year of the before period is estimated by the SPF relating traffic volume and other characteristics.

$$m = w_1(x) + \mathrm{w}_2(P) \tag{4}$$

where,
m : the number of expected crashes before treatment,
x : the number of crashes in the before period at a treatment site, and
P : the sum of annual SPF estimates.

From the mean and variance of the SPF estimate, the w_1 and w_2 are estimated. Where k estimated from the process of SPF estimation is a constant.

$$w_1 = \frac{P}{(P + \frac{1}{k})} \tag{5}$$

$$w_2 = 1/k(P + \frac{1}{k}) \tag{6}$$

2.3 Mean Absolute Deviation (MAD)

To assess the prediction performance of the models, the MAD is used (Lu *et al*., 2012). The MAD is defined:

$$MAD = \frac{1}{n}\sum \left| N_{observed_i} - N_{predicted_i} \right| \tag{7}$$

where,
n : the number of segments in the prediction dataset,
$N_{observed_i}$: observed crash frequency for segment i, and

$N_{predicted_i}$: predicted crash frequency for segment i.

3. DATA DISCRIPTION

In this study, six years of crash data from 2005 to 2010 were extracted from three Korean expressways operated by Korea Expressway Corporation. The four Korean expressways used in this study are Gyeongbu, Seohaean and Honam. The freeway segments were divided into interchanges or junctions, so the traffic volume is constant at each segment.

The segments comprise a total of 949.9km, and a total of 5,693 crashes occurred over six years on these expressways. The AADT and segment length were used as the exposure variables. Figure 1 shows the four Korean expressways used in this study and Table 1 shows the statistics of AADT, segment length, the number of crashes of each expressway.

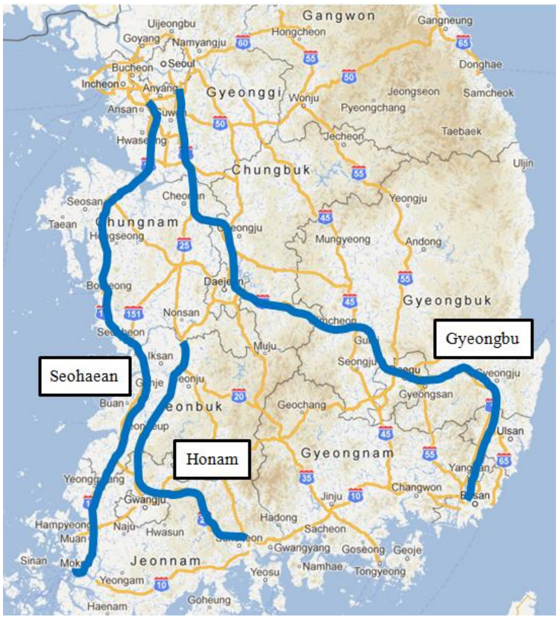

Figure 1. Three Korean expressways

Table 1. Data statistics of variables

Variable	Descriptive Statistics	Year					
		2005	2006	2007	2008	2009	2010
Number of total crashes	Minimum value	0	0	0	0	0	0
	Maximum value	27	22	21	19	18	21
	Average value	4.95	4.55	4.37	4.06	3.62	3.41
	Standard deviation	4.42	4.22	3.72	3.48	3.24	3.17
AADT (vehicles/year)	Minimum value	3,773	4,031	4,222	3,382	4,449	4,739
	Maximum value	106,213	108,152	111,356	100,132	99,512	98,755
	Average value	30,892	31,071	32,085	30,739	31,363	33,063
	Standard deviation	22,374	22,809	23,730	22,383	22,106	22,471
Segment Length (km)	Minimum value	0.6					
	Maximum value	28					
	Average value	8.45					
	Standard deviation	5.25					

4. ANALYSIS and RESULT

4.1 Predictability Validation

The data were divided into two groups: SPF estimation and predictability validation. Three years of crash data from 2005 to 2007 were used to estimate SPFs, and remaining data from 2008 to 2010 were used to observe the predictability of the SPFs.

The SPFs were estimated by linear regression, Poisson regression, and NB regression using the crash data from 2005 to 2007. The coefficients of SPF by linear, Poisson, NB regression are shown in Table 2. In case of possion and NB regression, four functional forms were estimated.

- Form A (Linear with AADT and L) : $Y = \alpha + \beta_1 AADT + \beta_2 L$
- Form B (Exponential with AADT and L) : $Y = e^{\alpha + \beta_1 AADT + \beta_2 L}$
- Form C (Power with AADT, Linear with L) : $Y = e^{\alpha} \times AADT^{\beta_1} \times L$
- Form D (Power with AADT, L) : $Y = e^{\alpha} \times AADT^{\beta_1} \times L^{\beta_2}$

Table 2. Coefficients of SPFs for the predictability validation

Regression	Functional Form	Coefficient					
		α		β_1		β_2	
		Estimate	P-value	Estimate	P-value	Estimate	P-value
Linear	A	-1.209	.000	0.000059	.000	0.471	.000
Poisson	B	0.411	.000	0.000011	.000	0.080	.000
	C	-5.007	.000	0.443389	.000	-	-
	D	-3.917	.000	0.367626	.000	0.830	.000
NB (over-dispersion parameter)	B (0.280)	0.318	.000	0.000011	.000	0.089	.000
	C (≒0)	-5.007	.000	0.443390	.000	-	-
	D (0.245)	-3.833	.000	0.360829	.000	0.824	.000

On the basis of estimating SPFs using the crash data from 2005 to 2007, the predicted values from 2008 to 2010 were estimated to observe the predictability. Then, the MAD was calculated by difference between predicted and observed value from 2008 to 2010. The MAD of SPFs is shown in Table 3.

Table 3. MAD of SPFs for the predictability validation

Regression	Functional Form	Mean Absolute Deviation (SD)
Linear	A	2.053 (1.65)
Poisson	B	2.033 (1.86)
	C	2.018 (1.68)
	D	2.009 (1.61)
NB	B	2.046 (1.81)
	C	2.018 (1.68)
	D	2.011 (1.60)
Empirical Bayes Estimation	B	1.880 (1.64)
	C	2.018 (1.68)
	D	1.472 (1.64)

4.2 Cross Validation

The data were divided into two groups: SPF estimation and cross validation. The half of crash data from 2008 to 2010 were used to estimate SPFs, and the cross validation was practiced by calculating the MAD between predicted and observed value with the remaining crash data from 2008 to 2010.

The SPFs were estimated by linear regression, Poisson regression, and NB regression using the half of crash data from 2008 to 2010. The coefficients of SPF by linear, Poisson, NB regression are shown in Table 4. In case of possion and NB regression, four functional forms were estimated.

Table 4. Coefficients of SPFs for the cross validation

Regression	Functional Form	Coefficient					
		α		β_1		β_2	
		Estimate	P-value	Estimate	P-value	Estimate	P-value
Linear	A	-1.499	.000	0.000058	.000	0.424	.000
Poisson	B	0.145	.050	0.000012	.000	0.084	.000
	C	-6.475	.000	0.572324	.000	-	-
	D	-5.216	.000	0.480058	.000	0.817	.000
NB (over-dispersion parameter)	B (0.174)	0.118	.246	0.000013	.000	0.085	.000
	C (≒0)	-6.475	.000	0.572324	.000	-	-
	D (0.490)	-5.177	.000	0.485338	.000	0.773	.000

After estimating SPFs using the half of crash data from 2008 to 2010, the predicted values from 2008 to 2010 were estimated using the remaining data. Then, the cross validation was practiced by calculating the MAD between predicted and observed value with the remaining crash data from 2008 to 2010. The MAD of SPFs is shown in Table 5.

Table 5. MAD of SPFs for the cross validation

Regression	Functional Form	Mean Absolute Deviation (SD)
Linear	A	2.174 (1.68)
Poisson	B	2.175 (1.20)
	C	2.048 (1.81)
	D	1.952 (1.64)
NB	B	2.180 (1.98)
	C	2.048 (1.81)
	D	1.956 (1.65)
Empirical Bayes Estimation	B	0.668 (0.48)
	C	2.048 (1.81)
	D	0.237 (0.24)

5. CONCLUSIONS

To investigate the validity of the safety performance function in Korean expressways, the predictability validation and the cross validation were employed. Four functional forms, linear regression, Poisson regression, NB regression and empirical Bayes estimation were used to compare the prediction performances. The predictability validation and the cross validation were practiced by using the mean absolute deviation (MAD).

In case of functional form C, the over-dispersion parameter was close to zero. The MADs were same for the Poisson, NB regression and empirical Bayes estimation for both validations. Analyzing among distributions of disturbance and various functions formulated included the empirical Bayes estimation, the MADs by linear, Poisson and NB regressions were very similar. However, the MADs by empirical Bayes estimation were lower than those of other regressions except for functional form C. On the other hand, analyzing among relationship between the number of crashes and exposures, the MADs by power function with AADT and L were lower than others. It means that the power function with exposures fit well and the results of the empirical Bayes estimation were the most satisfying.

REFERENCES

American Association of State Highway and Transportation Officials (AASHTO) (2010) *Highway Safety Manual*. AASHTO, Washington, D.C.

Ben-Akiva, M., Lerman, S.R. (1991) *Discrete Choice Analysis*. MIT Press, Cambridge.

Elvik, R. (1997) Effects on accidents of automatic speed enforcement in Norway. *Transportation Research Record*, 1595, 14-19.

Federal Highway Administration (FHWA) Crash Modification Factor Clearinghouse. http://www.cmfclearinghouse.org/. Accessed on February 25, 2013.

Griffin, L.I., Reinhardt, R.N. (1996) *A review of Two Innovative Pavement Patterns that Have Been Developed to Reduce Traffic Speeds and Crashes*. AAA Foundation for Traffic Safety, Washington, D.C.

Gross, F., Peransud, B.N., Lyon, C. (2010) *A Guide to Developing Quality Crash Modification Factors*. FHWA-SA-10-032, FHWA, U.S. Department of Transportation.

Hauer, E. (1997) *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon, Oxford, U.K.

Hauer, E., Harwood, D.W., Council, F.M, Griffith, M.S. (2002) Estimating safety by the empirical Bayes method: A tutorial. *Transportation Research Record*,1784, 126-131.

Hauer, E., Bonnesson, J.A., Council, F.M., Srinivasan, R., Zegeer, C. (2012) Crash modification factors: Foundational issues. *Transportation Research Record*, 2279, 67-74.

Park, E.S., Park, J., Lomax, T.J. (2010) A fully Bayesian multivariate approach to before–after safety evaluation. *Accident Analysis and Prevention*, 42, 1118-1127.

Persaud, B.N., Lyon, C. (2007) Empirical Bayes before-after safety studies: Lessons learned from two decades of experience and future directions. *Accident Analysis and Prevention*, 39, 546-555.

Persaud, B.N., Lan, B., Lyon, C., Ghim, R. (2010) Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. *Accident Analysis and Prevention*, 42, 38-43.

Tribbet, L., McGowen, P., Mounce, J. (2000) *An Evaluation of Dynamic Curve Warning Systems in the Sacramento River Canyon*. California Department of Transportation, Sacramento.

Tsyganov, A.R., Warrenchuk, N.M., Machemehl, R.B. (2009) Driver performance and safety effects of edge lines on rural two-lane highways. 88th Annual Meeting of Transportation Research Board, Washington, D.C., USA.

Washington, S.P., Karlaftis, M.G., Mannering, F.L. (2003) *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall, Boca Raton.