

## **Behind Traffic Fatality Reduction in 2008 in Virginia**

Young-Jun Kweon, Ph.D.  
Virginia Department of Transportation  
Email: Young-Jun.Kweon@VDOT.Virginia.Gov  
Phone: 434-293-1949

Cheryl W. Lynn  
Virginia Department of Transportation  
Email: Cheryl.Lynn@VDOT.Virginia.Gov  
Phone: 434-293-1905

### **ABSTRACT**

In 2008, the U.S. witnessed 9% reduction in traffic fatalities across the nation. In Virginia, 205 less persons were killed in traffic crashes in 2008 than in 2007, which is an unprecedented one-year decline since 1950, following by 170 reduction in 1974 allegedly attributable to oil crisis. Part of the reduction can probably be ascribed to continuous efforts for safety improvements in education, law, policy, technology, and vehicle and road design. However, some portion of the reduction may be attributable to changes in economic conditions in late 2007 and 2008. In this regard, this study attempted to find contributing factors to the 2008 reduction in traffic fatalities using Virginia data, and economic indicators (e.g., unemployment rate), crash exposures (e.g., vehicle miles traveled), and other factors (e.g., beer consumption) were attempted to relate to changes in fatalities using a time-series model. Changes in total employment number and unemployment rate were determined to be the best contributing factors for changes in annual traffic fatalities. However, only the unemployment rate was able to explain a portion of the reduction in 2008, 27%.

## INTRODUCTION

In 2008, all states in the United States except for four states (Delaware, New Hampshire, Vermont, and Wyoming) witnessed reduction in traffic fatalities, ranging from 1% reduction in Colorado up to 24% reduction in Alaska (National Highway Safety Administration [NHTSA], 2009). A 9.7% reduction (about 4,000 less fatalities) was recorded for the entire nation and 20% reduction (203 less fatalities) was recorded in Virginia. In the United States, there was also 5.8% reduction in traffic injuries and 3.5% reduction in total traffic crashes.

Since the reduction was nationwide in 2008, the extent of influences of contributing factors should be nationwide. For example, there was 3.4% nationwide reduction in VMT that is quite matched with 3.4% reduction in total traffic crashes (NHTSA, 2009). Economic activities are believed to derive transportation activities. Thus, it was speculated that economic changes from late 2007 through 2008 affect economic activities, influencing travel patterns and driving behaviors that are closely related to traffic safety. Thus, economic factors were considered as potential contributing factors to the traffic fatality reduction in 2008.

In this aspect, this study attempted to find factors being linked to reduction in the traffic fatalities in 2008. Finding contributing factors would be also useful for assessing and, if necessary, revising the 2010 reduction target in traffic fatalities set for Virginia's Strategic Highway Safety Plan Surface Transportation Safety Executive Committee, 2006). This study used time-series data of Virginia and utilized regression analysis methods to find factors contributing to the 2008 reduction in fatalities in Virginia. Since the study began in the first quarter of 2009, data from 2008 were limited. Limitations on data availability resulted in the consideration of only fatality data since injury data was not yet available.

## DATA

Empirical data were collected on the potential contributing factors identified by literature review. For some of the factors, data aggregated for different intervals were found. For example of Gross Value-weighted Industrial Production (GVIP), two-interval data, quarterly as well as yearly values, were found to be available. For consistent data analysis in terms of data intervals, one data interval was recommended. Among the three commonly found intervals, month, quarter, and year, the year interval was selected for this study for two reasons. First, the reduction goal in the Virginia's Strategic Highway Safety Plan was set on an annual basis (i.e., one year interval). Second, if an interval less than one year was adopted, outputs (e.g., prediction) generated from analysis of data aggregated for the interval are also at the same interval, which is less than one year. Thus, an accumulation of the outputs to one year is needed to suggest an annual reduction goal. Such an accumulation is likely to introduce additional complication (e.g., estimation of errors of the annual accumulated outputs) to analysis. In the end, empirical data of the identified potential contributing factors were collected on an annual basis for this study.

Not all the identified factors were available on an annual basis. Table 1 describes data collected on an annual basis, and two types of contributing factors are found, traffic crash exposure (e.g., vehicle miles traveled) and socio-demographic-economic factors (i.e., unemployment rate). Although there were five crash variables collected, only fatality data were available for 2008 and were used in this study. Thus, all analyses were performed to link fatality counts (dependent variable) to the contributing factors (explanatory variables).

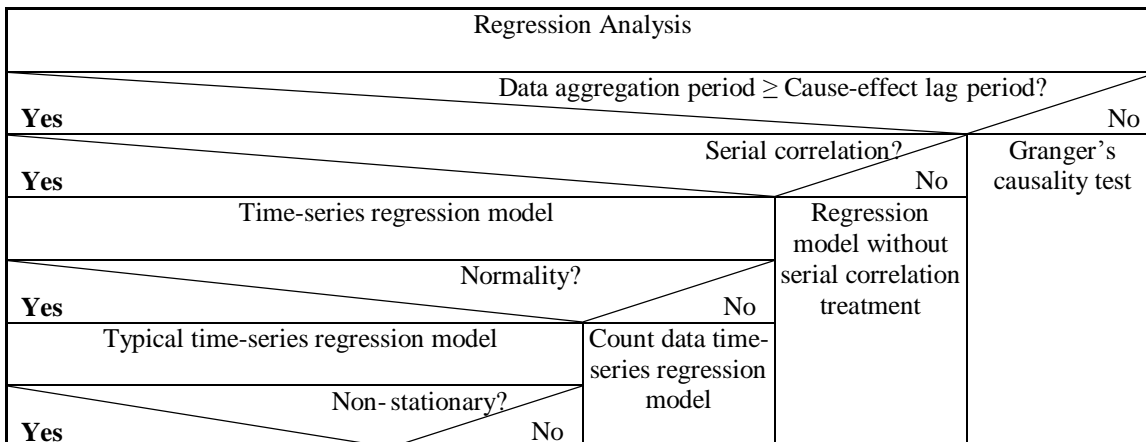
Table 1. List of Variables Collected for the Study

Type	Variable Name	Description	Data Period
Dependent Variables	Fatality	Number of fatalities	1951-2008
	Fatal Crash	Number of fatal crashes	1969-2007
	Injury	Number of injuries	1951-2007
	Injury Crash	Number of injury crashes	1969-2007
	Crash	Number of total crashes	1951-2007
Explanatory Variables	Beer	Per Capita Beer Consumption	1994-2008
	CAP	Manufacturing Capacity Utilization Base 2002	1967-2008
	CPI_UC4	US Consumer Price Index (All Urban Consumers)	1976-2008
	Driver	Number of registered drivers	1971-2007
	Employment	Employment	1976-2008
	Gas Price	US Gas Price (cents)	1990-2008
	Gas Production	Gasoline Through Company Outlets Volume by Refiners (Thousand Gallons per Day)	1985-2008
	GVIP	Gross Value-weighted Industrial Production Base 2000	1972-2008
	Income	Per Capital Disposable Personal Income	1958-2008
	IPSIC	Industrial Production: Manufacturing (SIC) Base 2000	1951-2008
	Labor Force	Labor Force	1976-2008
	Population	Population	1968-2008
	PPI_Com4	U.S. Producer Price Index (Commodity)	1976-2008
	Sales	Taxable Sales	1985-2008
	Unemployment	Unemployment	1976-2008
	Vehicle	Number of registered vehicles	1951-2007
VMT	Total vehicle miles traveled	1951-2007	

**METHOD**

**Regression Analysis**

A Literature review was performed to identify suitable methods for analyzing yearly data. A regression analysis was selected for this study because it can relate a dependent variable of interest (i.e., the annual number of fatalities) to multiple independent variables and result in an equation that reflects the relationship among variables and that can be used to revise the 2010 reduction target in traffic fatalities. In a regression analysis, there are various regression models available. Figure 1 summarizes a procedure for selecting an appropriate regression model for the study. The steps selected for this study are indicated by bold font.



<b>Differencing</b>	No differencing			
<b>1. Raw number</b> <b>2. Percentage</b> <b>3. Logarithm of raw number</b>				

Figure 1. Flow chart for selecting an appropriate regression model for the study.

The above flow chart is explained in detail below with a separate section for each step.

**Data Aggregation Period** Granger's causality test (1969) was initially recommended for this study based on research by Hoxie et al. (1984). However, it was discovered that the test would probably be invalid if data were aggregated over a period longer than a lag period of a cause-effect phenomenon. Other studies warned about distortion of the test results when data aggregation is involved (e.g., Gulasekaran, and Abeysinghe, 2002; Maminggi, 1996). For example, if an effect was realized one month after a cause occurred, data should not be aggregated over longer than a month (e.g., quarterly or yearly data), negating the use of Granger's causality test. Using data aggregated over a period longer than the lag period would lead to failure to discover a cause-effect relationship among the data, if the relationship exists.

Because the study data were prepared on an annual basis, any cause-effect relationship that might exist with a less than one year lag period would probably not be uncovered regardless of analysis methods. It was conjectured that, if a change in values of any of the explanatory variables in Table 1 caused a change in the number of fatalities, the effect due to the cause would probably be realized in a period shorter than the aggregation interval (i.e., 1 year). For example, we cannot expect that an increase in employment in January would influence a change in traffic fatalities in January of the next year. Thus, without data aggregated over a period much shorter than a year (e.g., monthly data), Granger's causality test is not applicable.

**Serial Correlation** Since the data are time-series in nature, a serial correlation may well exist and, if so, it would interfere with a statistical significance of the relationship. Thus, a statistical test for serial correlations should be performed to detect the presence of serial correlations. If serial correlations are found, a regression model that corrects for the correlations inherent in time-series data should be used. In this study, serial correlations were found in the numbers of fatalities over years using the Durbin-Watson test. Thus, a regression model with correction for serial correlations was employed.

**Normality** Normality should also be checked on a dependent variable used in the regression model. When the distribution of the variable is found to severely depart from a normal distribution, a model designed for non-normally distributed variables should be employed. For count time-series data, an Integer-valued Auto-Regressive (INAR) model becomes the most common application in recent years. The dependent variable of this study, the number of fatalities, was found to be normally distributed based on the Shapiro-Wilk test that is generally considered to be robust for sample sizes less than 2,000. Thus, a time-series model with a normal distribution assumption was employed.

**Non-Stationarity and Differencing** In applying a typical time-series model, it is assumed that variables are stationary (i.e., mean and variance are constant over time). When the variables are non-stationary (especially for the mean over time), differencing generally removes non-stationarity from the data series. For this study, non-stationarity was found for almost all the

variables in Table 1 based on Phillips-Perron unit root test. Thus, first differencing was applied to convert the non-stationary variables to stationary variables. Moreover, year-to-year changes in raw numbers or percentages might be more appropriate to examine for contributing factors in a time-series setting. This study, in fact, examined three variants of differenced variables – difference in raw values, percentages, and logarithm of raw values.

### Autoregressive Error Model

An autoregressive error model has been successfully employed in the past (e.g., Kweon, 2006 and 2007; Raeside, 2004; Oppe, 1989) and was employed in this study. Two sets of autoregressive error models were developed, one with lagged fatality variables and the other without them – lagged variables are variables with values in previous time periods. The two sets of autoregressive error models for the three differencing variants can be expressed mathematically as follow:

#### *Models with lagged fatality variables*

$$\Delta Y_t = \alpha + \beta_1 \times \Delta Y_{t-1} + \beta_2 \times \Delta Y_{t-2} + \dots + \gamma \times \Delta X_t + v_t \text{ (change in raw values),}$$

$$\Delta \log Y_t = \alpha + \beta_1 \times \Delta Y_{t-1} + \beta_2 \times \Delta Y_{t-2} + \dots + \gamma \times \Delta \log X_t + v_t \text{ (change in logarithm of raw values),}$$

$$\% \Delta Y_t = \alpha + \beta_1 \times \% \Delta Y_{t-1} + \beta_2 \times \% \Delta Y_{t-2} + \dots + \gamma \times \% \Delta X_t + v_t \text{ (percentage change in raw values), and}$$

$$v_t = \phi_1 v_{t-1} + \phi_2 v_{t-2} + \dots + \phi_p v_{t-p} + \varepsilon_t, \varepsilon_t \sim IN(0, \sigma^2) \quad \text{Eq. (2)}$$

#### *Models without lagged fatality variables*

$$\Delta Y_t = \alpha + \gamma \times \Delta X_t + v_t \text{ (change in raw values),}$$

$$\Delta \log Y_t = \alpha + \gamma \times \Delta \log X_t + v_t \text{ (change in logarithm of raw values),}$$

$$\% \Delta Y_t = \alpha + \gamma \times \% \Delta X_t + v_t \text{ (percentage change in raw values), and}$$

$$v_t = \phi_1 v_{t-1} + \phi_2 v_{t-2} + \dots + \phi_p v_{t-p} + \varepsilon_t, \varepsilon_t \sim IN(0, \sigma^2) \quad \text{Eq. (3)}$$

where  $\Delta$  is a year-to-year difference operator (e.g.,  $\Delta Y_t = Y_t - Y_{t-1}$  and

$$\% \Delta Y_t = 100 \times (Y_t - Y_{t-1}) / Y_t),$$

$Y_t$  is a dependent variable (i.e., the number of fatalities) in year  $t$ ,

$X_t$  is an independent variable, which is listed in Table 1, in year  $t$ ,

$v_t$  is an correlated error term at time  $t$ , and

$\phi_1, \dots, \phi_p$  are autoregressive parameters.

A model was developed with one factor analyzed at a time. Factors that were statistically significant at 0.05 level were determined to be contributing to changes in fatalities. A model with the highest coefficient of determination (i.e., R-squared measure) was considered to be the best model and the factor included in that model was considered to be the best contributing factor. In cases where more than one contributing factor was found, factors other than the best contributing factor were entered into the best model when the independent variables were not significantly correlated with one another (e.g., correlation coefficient  $\geq 0.7$ ). When added factors

are statistically related to the dependent variable at 0.05 level, the model with the added factors along with the best factor becomes the new best model.

## RESULTS AND DISCUSSION

### Regression Analysis

The results of the preliminary data analysis provided some guidance on proper forms of variables for regression analysis. Specifically, using raw values of the variables was not likely to result in a statistically significant relationship between fatality and the possible factors. Instead, year-to-year difference values, called first differencing, were considered for regression modeling to find such a relationship. Also, use of differenced variables might be more appropriate in finding contributing factors in time-series data. Moreover, non-stationarity was found for almost all the variables used in the study based on Phillips-Perron unit root test, which requires first differencing of the series to convert to stationarity. With these reasons, year-to-year differenced values instead of raw values were used for developing final regression models.

In addition to the first differencing in raw values, the first differencing in percentages and logarithm of raw values were also employed. Thus, three different differenced variables were used as a dependent variable as presented in Eqs. (2) and (3). However, results using these three variables were generally in agreement with one another. Thus, this study reports results only from analysis of the differenced variable in raw values. Table 2 summarizes results of regression analysis.

Table 2. Association of Factors to Fatality in terms of Change in Raw Values

Variable	Association with $\Delta$ Fatality <sup>1</sup>		Coefficient of Determination <sup>2</sup>	
	Model w/ <sup>3</sup>	Model w/o <sup>4</sup>	Model w/ <sup>3</sup>	Model w/o <sup>4</sup>
$\Delta$ Beer	No	No	–	
$\Delta$ CAP	No	No	–	
$\Delta$ CPI_UC4	No	No	–	
$\Delta$ Driver	No	No	–	
<b><math>\Delta</math>Employment</b>	Yes	Yes	<b>0.3099</b>	0.1774
$\Delta$ Gas Price	No	No	–	
$\Delta$ Gas Production	No	No	–	
$\Delta$ GVIP	Yes	Yes	0.2456	0.1624
$\Delta$ Income	No	No	–	
$\Delta$ IPSIC	Yes	Yes	0.2504	0.1603
$\Delta$ Labor Force	No	No	–	
$\Delta$ Population	No	No	–	
$\Delta$ PPI_Com4	No	No	–	
$\Delta$ Sales	No	No	–	
$\Delta$ Unemployment	Yes	Yes	0.3089	0.2300
<b><math>\Delta</math>Unemployment Rate</b>	Yes	Yes	0.3053	<b>0.2306</b>
$\Delta$ Vehicle	No	No	–	
$\Delta$ VMT	No	No	–	

<sup>1</sup>Yes if it is statistically significant at 0.05 level (95% confidence level); No otherwise

<sup>2</sup>R-squared value (presented only for the factors with statistically significant association with  $\Delta$ Fatality)

<sup>3</sup>Model with lagged fatality variables

<sup>4</sup>Model without lagged fatality variables

According to the coefficient of determination in Table 2, for the model with lagged fatality variables, employment is the best contributing factor while GVIP, IPSIC, unemployment number, and unemployment rate are contributing factors. GVIP and IPSIC was added, one at a time, to the model with the best contributing factor (employment) yet it became statistically insignificant: since GVIP and IPSIC are highly correlated, both could not enter in the model at the same time. Unemployment number and rate could not enter into the model because of their high correlations with employment. Analysis using two other differenced variables, difference in logarithmic values and percentage change, also arrived at the same conclusion as above.

For the model without lagged fatality variables, an unemployment rate is the best contributing factor while employment, GVIP, IPSIC, and unemployment number are contributing factors. GVIP and IPSIC was added, one at a time, to the model with the best contributing factor (unemployment rate) yet it became statistically insignificant. The unemployment rate could not enter into the model due to its high correlation with the employment number. The two final models with the best contributing factors are presented below separately.

**Model With Lagged Fatality Variables** The model with the best contributing factor, employment, was estimated as follows:

$$\Delta Fatality_t = -67.8734 - 0.3828 \times \Delta Fatality_{t-1} + 0.001054 \times \Delta Employment_t \quad \text{Eq. (4)}$$

All parameter estimates are statistically significant at 0.05 level and Durban-Watson test confirmed that there is no remaining serial correlation in regression residuals. The model can be expressed in the original raw number format as follows:

$$Fatality_t = -67.8734 + 0.6172 \times Fatality_{t-1} + 0.3828 \times Fatality_{t-2} + 0.001054 \times (Employment_t - Employment_{t-1}) \quad \text{Eq. (5)}$$

The coefficient estimate of the change in employment, 0.001054, can be translated that every 1,000 job lost a reduction of one fatality can be expected. When applying this finding on fatalities in 2008, an increase of 14,532 in employment recorded from 2007 to 2008 is predicted to increase traffic fatalities by 15. Thus, the employment change in 2008 does not explain fatality reduction in 2008.

**Model Without Lagged Fatality Variables** The model with the best contributing factor, unemployment rate, was estimated as follows:

$$\Delta Fatality_t = -9.4426 - 55.5130 \times \Delta Unemployment Rate_t \quad \text{Eq. (6)}$$

All parameter estimates are statistically significant at 0.05 level and Durban-Watson test confirmed that there is no remaining serial correlation in regression residuals. The model can be expressed in the original raw number format as follows:

$$Fatality_t = -9.4426 + Fatality_{t-1} - 55.5130 \times (Unemployment Rate_t - Unemployment Rate_{t-1}) \quad \text{Eq. (7)}$$

The coefficient estimate of the change in unemployment rate, -55.5130, can be translated such that every percentage increase in unemployment rate is associated with reduction of about 55 fatalities. When applying this finding on fatality in 2008, an increase of 1% in unemployment rate observed from 2007 to 2008 is predicted to reduce traffic fatality by 55 persons. Unemployment rate change in 2008 does explain a portion (27%) of the large fatality reduction in 2008.

## FINDINGS AND CONCLUSIONS

This study attempted to link annual traffic fatality fluctuation to annual changes in 17 potential contributing factors using Virginia data and time-series regression models with an anticipation of finding factors contributing to the reduction in traffic fatalities in 2008. Based on the analysis results, the following conclusions are drawn:

- Changes in values from year to year were found to result in better model results than the raw numbers themselves did.
- The employment number and unemployment rate are the two best contributing factors to traffic fatalities in terms of annual changes, among the 17 potential factors considered in this study.
- Other contributing factors are found to be the Gross Value-weighted Industrial Production (GVIP) and the Industrial Production Index of Manufacturing. Thus, all the contributing factors found in this study are economic indicators.
- As for the fatality reduction observed in 2008 (i.e., 205), the employment number change in 2008 failed to explain the reduction while the unemployment rate change in 2008 explained a portion of the entire reduction, 27%.

One notable limitation exists in this study. Using annual data prohibited from performing statistical tests of causality between traffic fatality and contributing factors. Monthly data should be utilized so that a cause-effect relationship can be statistically tested.

## ACKNOWLEDGEMENT

The authors thank members of Virginia's Surface Transportation Safety Committee for their reviews and comments on the study and Virginia Department of Transportation for sponsoring the study.

## REFERENCES

- Granger, C.W.J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37, no. 3 (1969): 424-438.
- Gulasekaran, R. and T. Abeyasinghe. The Distortionary Effects of Temporal Aggregation on Granger Causality. Working Paper No. 0204, Department of Economics, National University of Singapore, May 2002. Available at [www.fas.nus.edu.sg/ecs/pub/wp/wp0204.pdf](http://www.fas.nus.edu.sg/ecs/pub/wp/wp0204.pdf). Accessed on August 31, 2009.
- Hoxie, P., D. Skinner, and G. H. Wang. Socio-Economic Influences on Highway Fatalities: An Empirical Investigation. Report DOT-HS-806-525. U.S. Department of Transportation, Washington, D.C., 1984.
- Kweon, Y.-J. Evaluation of a Proposal to Set a Goal for the Virginia Strategic Highway Safety Plan of a Forty Percent Reduction in Traffic Fatalities and Injuries by 2010. Final Report VTRC 06-R34, Virginia Transportation Research Council, Charlottesville, Virginia, June 2006.
- Kweon, Y.-J. "Forecasting Traffic Safety in Virginia with a Primary Seat Belt Law." In CD-ROM of ITE 2007 Annual Meeting and Exhibit Compendium of Technical Papers, Pittsburgh, Pennsylvania, August 2007.

- Maminggi, N. "Aggregation over Time, Error Correction Models and Granger Causality: A Monte Carlo Investigation." Economics Letters 52, no. 1 (1996): 7-14.
- National Highway Traffic Safety Administration (NHTSA). Traffic Safety Facts: Crash Stats. Publication DOT-HS-811-172, U.S. Department of Transportation, June 2009.
- Oppe, S. "Macroscopic Models for Traffic and Traffic Safety." Accident Analysis and Prevention 21, no. 3 (1989): 225-232.
- Raeseide, R. "Predicting and Monitoring Casualty Numbers in Great Britain." Journal of Transportation and Statistics 7, no. 1 (2004): 61-68.
- Surface Transportation Safety Executive Committee. Commonwealth of Virginia's Strategic Highway Safety Plan 2006-2010. Richmond, Virginia, 2006.