

A Study on Kentucky Highway Fatalities

Ryan Glassford, University of Louisville, Louisville, KY

ABSTRACT

OBJECTIVE: To attempt to forecast the number of highway fatalities each month in the state of Kentucky, using rain occurrences as a dynamic regressor. **METHODS:** The data used is a month-by-month list of fatalities from 2000-2005 occurring in Kentucky. The data were compiled from the Fatality Analysis Reporting System (FARS), a database provided by the U.S. Department of Transportation. SAS version 9.1 is used to predict an ARIMA model with fatalities as the dependent factor and time and rainfall as independent factors. **RESULTS:** The best ARIMA model to fit seems to be $ARIMA(1,0,2) \times (0,0,1)_{12}$ in conjunction with the dynamic regressor model of rainfall deaths. The parameters for this model are a lag of 3, and simple numerator and denominator factors of 1 and 3, respectively. This combination fits the data well in the hold-out sample of 2005, and the overall mean average percent error is 4.76. **CONCLUSIONS:** In retrospect, it appears that the number of highway fatalities is seasonal and has an increasing trend. It also appears that the late summer months have the highest incidence of fatalities, as opposed to the early spring months--using rainfall deaths as a dynamic regressor helps minimize these effects in the model.

INTRODUCTION

Getting from place to place has been a necessity as long as people have been around, and doing so is much easier since the advent of automobiles and highway systems. However, there are also disadvantages. Thousands of people die every year in highway accidents across the world. In this study we examine accident fatalities in the state of Kentucky. The purpose is to construct a model with a dynamic regressor to estimate the distribution of deaths for each month: doing so could give important results that can be analyzed in the hopes of cutting down the number of fatalities in the future.

METHODS AND RESULTS

The dataset used was compiled from the Fatality Analysis Reporting System (FARS), a database compiled every year by the U.S. Department of Transportation and the National Traffic and Highway Safety Administration. The number of deaths each month was tabulated from January, 2000 to December, 2005, along with the proportion of these deaths resulting under the condition of rainfall. These data were then converted into a format readable by SAS 9.1 (© 2002-2003 SAS Institute Inc.).

The highway fatality data were then processed by the SAS Time Series Forecasting System (TSFS), where a suitable time series model was determined. Time in months is measured as our independent variable, used to forecast the total number of fatalities for a given time interval. The number accidents occurring during rainy weather is used as the dynamic regressor.

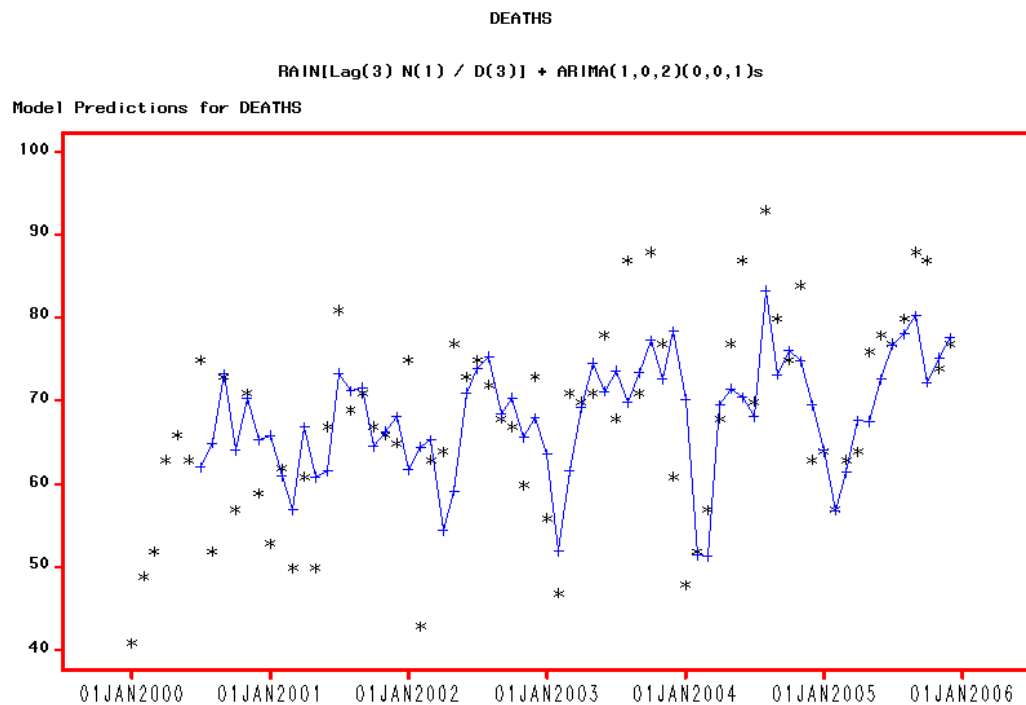
After running the TSFS, the highway data were fit to an $ARIMA(1,0,2) \times (0,0,1)_{12}$ model with no intercept. This model suggests an autoregressive factor of 1 and a moving average of 2; the use of differencing was discarded as the regressor was added. The model also proposes a seasonality factor of 12, with a moving average of 1 (Figure 1). This model is used in combination with the rainfall fatalities dynamic regressor. The regressor model uses a lag of 3, simple numerator of 1, simple denominator of 3, and no differencing or seasonal aspects.

FIGURE 1. MODEL FORECASTS FOR HIGHWAY DEATHS

pdfMachine - is a pdf writer that produces quality PDF files with ease!

Get yours now!

"Thank you very much! I can use Acrobat Distiller or the Acrobat PDFWriter but I consider your product a lot easier to use and much preferable to Adobe's" A.Sarras - USA



From Figure 1, we can see that the ARIMA model follows the actual values closely. It appears that the data have a small, positive trend, and an unidentifiable seasonal factor of around one year.

The statistics of fit were low, with MAPE being 4.76 and RMSE at 5.73. The holdout sample of the year of 2005 matched very closely with the predicted data, with the largest residual being 14.7 in October, 2005.

Another important graph to view is the autocorrelations (Figure 2).

FIGURE 2. AUTOCORRELATION PLOTS

pdfMachine - is a pdf writer that produces quality PDF files with ease!

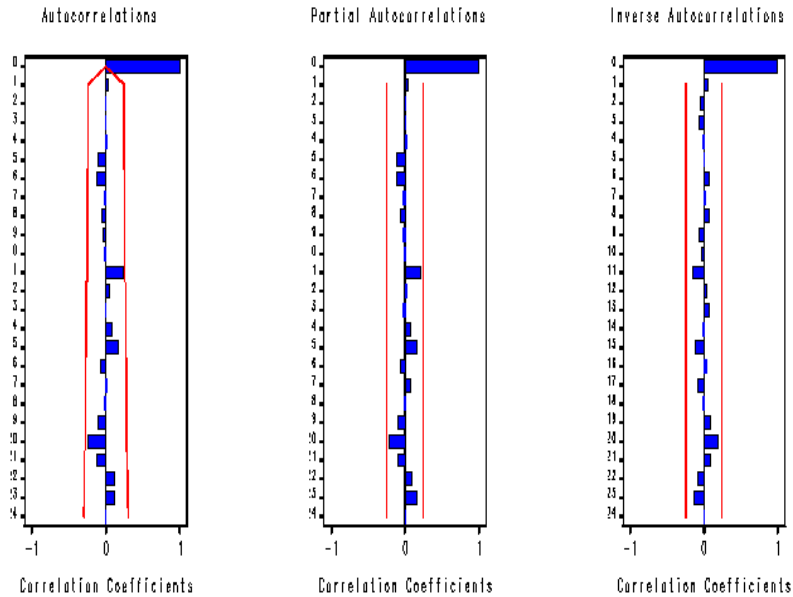
Get yours now!

"Thank you very much! I can use Acrobat Distiller or the Acrobat PDFWriter but I consider your product a lot easier to use and much preferable to Adobe's" A.Sarras - USA

Prediction Error Autocorrelation Plots

DEATHS

RAIN[Lag(3) N(1) / D(3)] + ARIMA(1,0,2)(0,0,1)s



We see that all the correlations fall into the zero confidence interval. This can be interpreted as being a good model of fit. Finally, we see that predictions of 2006 fit in with our model of an increase in fatalities. Figure 3 shows the 95% confidence interval of these estimates:

FIGURE 3. 2006 PREDICTIONS

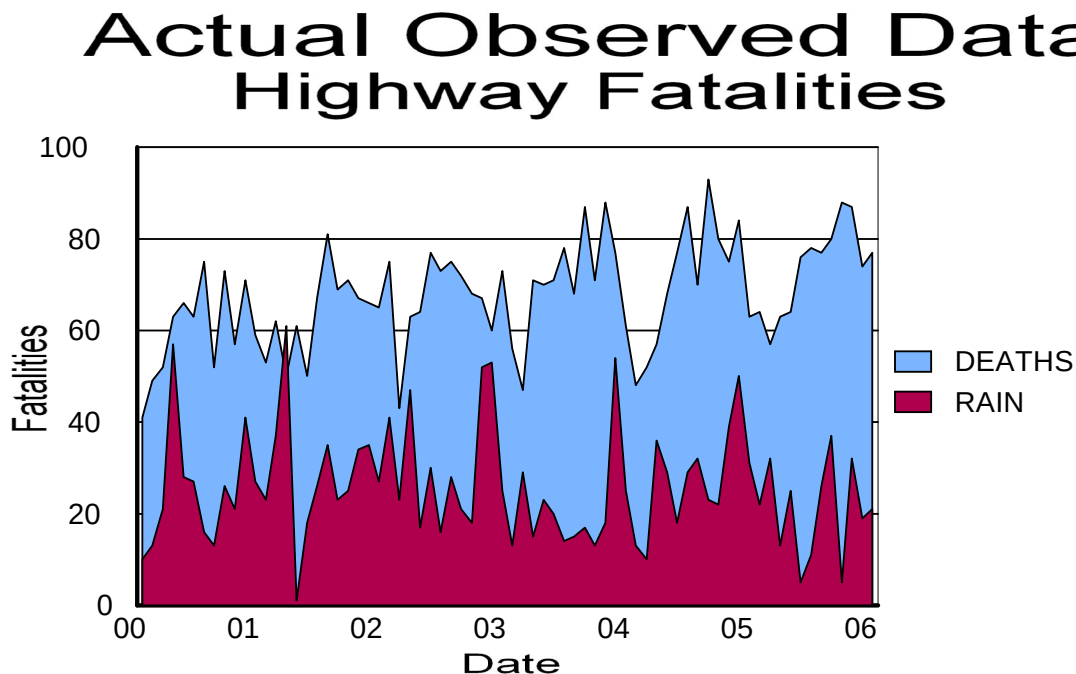
	Predicted	Upper 95%	Lower 95%
	value for	Confidence	Confidence
DATE	DEATHS	Limit	Limit
01JAN2006	74.3832	93.1717	55.5948
01FEB2006	65.7584	84.8050	46.7118
01MAR2006	72.8439	93.7060	51.9817
01APR2006	66.5695	87.6804	45.4586
01MAY2006	70.4723	91.6188	49.3258
01JUN2006	69.6006	90.7523	48.4489
01JUL2006	69.0042	90.1566	47.8518
01AUG2006	69.5968	90.7493	48.4443
01SEP2006	70.2757	91.4282	49.1231
01OCT2006	72.1309	93.2835	50.9784
01NOV2006	69.3742	90.5267	48.2216
01DEC2006	69.8346	90.9872	48.6821

pdfMachine - is a pdf writer that produces quality PDF files with ease!

Get yours now!

"Thank you very much! I can use Acrobat Distiller or the Acrobat PDFWriter but I consider your product a lot easier to use and much preferable to Adobe's" A.Sarras - USA

FIGURE 4. ACTUAL OBSERVATIONS



CONCLUSION

There are obviously many limitations to this study. First of all, only six years of data were collected. Many factors may have changed before or after this sample, such as the safety of cars, an increase in inclement weather, and other human factors that are unaccounted for (the number of road miles). Also, this study was only for Kentucky, and in no means is intended for forecasting data in other parts of the world.

Further analysis of rainfall occurrences supplemented the ARIMA model as a dynamic regressor. As a result the MAPE was lowered by a value of almost three (7.20 without), and the model is generally a better fit than without this regressor. The predicted values are very close to the actual data. Analyses of the predictions reveal that the highest residuals occur when actual rainfall deaths are much lower than forecasted value during that given month.

The developed model did provide a good estimate for our hold out sample of 2005, with a low error percentage. The ARIMA models are a good way to try to forecast such data, and SAS provides a great environment to test such models.

Time series forecasting is a good tool to use in predicting factors such as highway safety. Many more variables would need to be introduced in order to find a more suitable model; this study only took into account time by month, total fatalities on Kentucky highways, and the proportion of deaths occurring during rainfall.

There may exist a better model to forecast these data. In subsequent studies, more variables will be added to the model in hopes of more accurately forecasting the data. In turn, authorities will be able to more accurately prepare for such events and possibly use preventative measures to avert future accidents, like for example enacting a law to lower the speed limit during periods of rain or other natural conditions. Such changes may prove to provide a safer highway environment and less fatal accidents.

CONTACT INFORMATION

pdfMachine - is a pdf writer that produces quality PDF files with ease!

Get yours now!

"Thank you very much! I can use Acrobat Distiller or the Acrobat PDFWriter but I consider your product a lot easier to use and much preferable to Adobe's" A.Sarras - USA

Your comments and questions are valued and encouraged. Contact the author at:

Ryan Glassford
University of Louisville
7413 Steeplecrest Circle #207
Louisville, KY 40222
(513) 608-8593
rjglas03@louisville.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

Other brand and product names are trademarks of their respective companies

pdfMachine - is a pdf writer that produces quality PDF files with ease!

Get yours now!

"Thank you very much! I can use Acrobat Distiller or the Acrobat PDFWriter but I consider your product a lot easier to use and much preferable to Adobe's" A.Sarras - USA