

**NBSIR 73-189**

# **A Comparison of Several Methods for Forecasting U.S. Traffic Fatalities**

---

Alexander R. Crow

Technical Analysis Division  
National Bureau of Standards  
Washington, D. C. 20234

May 1973

Final Report

Prepared for  
**National Highway Traffic Safety Administration**  
**Department of Transportation**  
Washington, D. C. 20590



NBSIR 73-189

**A COMPARISON OF SEVERAL METHODS FOR  
FORECASTING U.S. TRAFFIC FATALITIES**

---

Alexander R. Crow

Technical Analysis Division  
National Bureau of Standards  
Washington, D. C. 20234

May 1973

Final Report

Prepared for  
National Highway Traffic Safety Administration  
Department of Transportation  
Washington, D. C. 20590



---

**U. S. DEPARTMENT OF COMMERCE, Frederick B. Dent, Secretary**  
**NATIONAL BUREAU OF STANDARDS, Richard W. Roberts, Director**



## SUMMARY

This is a second report to the Mathematical Analysis Division of the National Highway Traffic Safety Administration (NHTSA) on the subject of forecasting annual highway fatalities. This report concerns a comparison of several time series analysis programs based on exponential smoothing and non-decompositional methods currently employed by NHTSA for projecting the annual traffic fatalities for the entire U.S. Several methods of data aggregation are studied.

It is found that there is some advantage in using lumped (pooled) data for each region aggregated either quarterly or half yearly, and using the Sum of Regional estimates to estimate the national value.

Also, there does not appear to be any great difference in the results obtained using the non-decompositional methods and those obtained by time-series analysis programs based on exponential smoothing methods.

Estimates for the 1972 and 1973 national traffic fatalities by a variety of methods were made. For 1973 the estimates ranged from a low of 54186 to a high of 55994, with a mean of 55055.

## TABLE OF CONTENTS

Acknowledgements

Summary

List of Tables

List of Symbols and Formulas

1. Introduction and Problem Definition
  2. Analyses Based on 12 Years and 6 Months of Data (12,6).
    - 2.1 National Traffic Fatalities
    - 2.2 A Look to Regions
  3. Forecasts of Total U.S. Traffic Fatalities Using No Data from Current Year
    - 3.1 Modifications Needed
    - 3.2 Extrapolations
  4. Conclusions
- Appendix: Comments on Forecasting Techniques and Computer Programs

LIST OF TABLES

- Table 1. Comparisons of Estimates by Data Features Based on  $\Delta_6/S^*$  (Fractional Error). Estimates for Calendar 1972, Using 12 Years, 6 Months Data
- Table 2. Errors of Estimates of 1972 U.S. Traffic Fatalities by Various Non-Decompositional Methods Using 12 Years, 6 Months of data.
- Table 3. Estimates of U.S. Traffic Fatalities for 1961-1972 Using a Difference Method:  $\hat{Y}_0 = m_0 + (Y_{-1} - m_{-1})$
- Table 4. Order of Region by Number of 1972 Traffic Fatalities
- Table 5. Tentative New Regions Based on Similarity of Seasonality Factors from TIMSR 4
- Table 6. Fractional Errors of Regional Estimates for 1972 by Various Methods Based on 12 Years, 6 Months of Data.
- Table 7. Frequency of Firsts, Seconds and Third Best Procedures (Regions)
- Table 8. Computation of Forecasts of U.S. Traffic Fatalities for 1973 by a Modified NHTSA Method:

$$\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{\sum_1^k Y_{-i+1}}{\sum_1^k m_{-i}} = \text{ratio of } k \text{ average pairs}$$

- Table 9. Computation of Estimates of U.S. Traffic Fatalities for 1971, 1972 by a Modified NHTSA Method:

$$\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{\sum_1^k Y_{-i+1}}{\sum_1^k m_{-i}}$$

- Table 10. Computation of Estimates of U.S. Traffic Fatalities for 1973, 1972 Using a Modified NHTSA Method:

$$\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{1}{k} \frac{\sum_1^k Y_{-i+1}}{m_{-i}} = \text{average of } k \text{ ratios}$$

Table 11. Computation of Estimates of U.S. Traffic Fatalities for 1969, 1970, 1971 Using a Modified NHTSA Method

$$\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{1}{k} \sum_{i=1}^k \frac{Y_{-i+1}}{m_{-i}} = \text{average of } k \text{ ratios}$$

Table 12. Computation of Estimates of U.S. Traffic Fatalities for 1962-1973 Using a Modified Difference Method.

$$\hat{Y}_{0+1} = Y_0 + (m_0 - m_{-1})$$

Table 13. Forecasts of U.S. Traffic Fatalities for 1973 Using a Variety of Methods

Table 14. Estimates of U.S. Traffic Fatalities for 1972 Using 12 Years of Data

Table 15. Fractional Errors of Regional Estimates for 1972 by Various Methods Based on 12 Years of Data.



## ACKNOWLEDGEMENTS

Particular thanks are due to Mr. William Obright, who made all of the computer runs and carried the major burden of the computing load. Mrs. Elizabeth Leyendecker passed along her knowledge of operating with the GEXSMO and EXPSMOOTHING programs and graciously offered counsel when snags appeared.

## LIST OF SYMBOLS AND FORMULAS

- $S$  = the actual (observed) value of the time series datum (in this case fatalities). This may be a monthly, quarterly, half-year or yearly value, depending on the context
- $S^*$  = the actual value for a calendar year
- $\hat{S}$  = an estimate of  $S$
- $\Delta$  =  $\hat{S} - S$ ; if  $\Delta < 0$ , the estimate is below the actual; if  $\Delta > 0$ , the estimate is above the actual
- $\Delta_6$  = the error in the forecast for the last 6 months of the calendar year (Table 1)
- $\Delta/S^*$  = the mean absolute fractional error
- $m_0$  = current total for specified months of fatalities from which calendar year forecasts are made. This may be for 3, 6 or 9 months. In this report a 6-month value is customary.
- $m_{-i}$  = the corresponding total for specified months of fatalities for  $i^{\text{th}}$  previous year
- $Y_{-i}$  = the total fatalities for the  $i^{\text{th}}$  previous year;  $i = 0, 1, 2, \dots$
- $\hat{Y}_j$  = an estimate of fatalities for year  $j$
- $K$  = the number of years of data over which summation occurs
- $N$  = an NHTSA forecast method
- $N_1^{(K)}$  = an NHTSA forecast method applied to the Lumping mode using  $K$  years of past data
- $N_2^{(K)}$  = an NHTSA forecast method applied to the Sum of States mode using  $K$  years of past data
- $M$  = monthly data
- $Q$  = quarterly data
- $H$  = half-year data
- $Y$  = yearly data

L(M) = a lumping mode (for a region or U.S.) using monthly data. Monthly data from individual states constituting a region are added together to create a time-series for a region

SS(Q) = a Sum of States mode consists of making individual analyses and predictions for each state, then sum the results over the states constituting a region, then sum the results of each of the regions in order to obtain a forecast for the natural values; in this instance quarterly data are indicated

$$\bar{L} = \frac{1}{4} [L(M) + L(Q) + L(H) + L(Y)]$$

= an average of lumping methods

$$\bar{L}_1 = \frac{1}{3} [L(Q) + L(H) + L(Y)]$$

} Regions = a forecast for the national value obtained by summing the individual forecasts for each region

Entire US = a forecast for the national value obtained from one grand lumped time series for the nation

$\lambda$  = a leap year adjustment factor

NHTSA FORECAST METHODS

$$N_1^{(K)}: \frac{\hat{Y}_0^{(K)}}{m_0} = \lambda \frac{\sum_1^K Y_{-i}}{\sum_1^K m_{-i}}, \text{ a ratio or proportionality method}$$

A Difference Method:

$$D_0: \hat{Y}_0 - m_0 = Y_{-1} - m_{-1} \rightarrow \hat{Y}_0 = m_0 + (Y_{-1} - m_{-1})$$

Modified NHTSA Forecast Methods:

$$M_1^{(K)}: \frac{\hat{Y}_{0+1}^{(K)}}{m_0} = \lambda \frac{\sum_1^K Y_{-i}}{\sum_1^K m_{-i+1}} = \text{ratio of } K \text{ average pairs, } K = 1, 2, \dots$$

$$M_2^{(K)}: \frac{Y_0^{(K)}}{m_0} = \frac{\lambda}{K} \sum_1^K \frac{Y_{-i+1}}{m_{-i+1}} = \text{average of } K \text{ ratios, } K = 1, 2, \dots$$

Modified Difference Methods:

$$D_1: Y_{0+1} - m_0 = Y_0 - m_{-1} \text{ or } Y_{0+1} = Y_0 + (m_0 - m_{-1})$$

Variants of this are:

$$D_1^{(K)}: \hat{Y}_{0+1}^{(K)} - m_0 = \frac{1}{K} \sum_{i=0}^{K-1} (Y_{-i} - m_{-i+1}) = \overline{(Y_0 - m_{-1})}^{(K)}, K = 1, 2, \dots$$

$$D_2^{(K)}: \hat{Y}_{0+1}^{(K)} = Y_0 + \frac{1}{K} \sum_0^{K-1} (m_{-i} - m_{-i+1}) = Y_0 + \overline{(m_0 - m_{-1})}^{(K)}$$

Note that in any of the methods with a superscript (K) will yield K estimates, say

$$\hat{Y}_0^{(1)}, \hat{Y}_0^{(2)}, \dots, \hat{Y}_0^{(K)}$$

and that these may be averaged to produce still other estimates.

In table headings:

GEXS  $\equiv$  GEXSMO time series analysis computer program

EXPS  $\equiv$  EXPSMOOTHING time series analysis computer program

MAE  $\equiv$  Mean absolute fractional error



## 1.0 INTRODUCTION AND PROBLEM DEFINITIONS

In a previous study<sup>1</sup> for the Mathematical Analysis Division (MAD) of the National Highway Traffic Safety Administration (NHTSA), the Technical Analysis Division made a comparison of forecasts for the current calendar year of national (entire U.S.) highway fatalities. Those obtained by using readily available computer routines for time series analyses based on exponential smoothing procedure were compared with methods employed currently by NHTSA. The methods were applied to estimate the current year total of fatalities given an initial 3 months of fatality data, then an initial 6 months of data and finally an initial 9 months of data. That report concluded, in part, that there is no coercive evidence to change their methods when operating on the national highway fatalities time series.

It was pointed out by members of NHTSA<sup>2</sup> that the cyclic and trend properties of the highway fatality time series appear to differ in the various regions and, perhaps, are idiosyncratic from state to state. Thus it appeared possible that some important information is lost when operating on the time series for the entire U.S. To explore this possibility it was suggested that forecasts for each region be made using an exponential smoothing method on

---

<sup>1</sup>U.S. Department of Commerce, National Bureau of Standards, Time Series Forecasting of Highway Accident Fatalities. NBSIR-73-138 (1973).

<sup>2</sup>Letter from R.J. Taylor, NHTSA, to Alexander Craw, NBS, January 29, 1973, Reference number N43-31.

each region's time series and that these forecasts be summed to obtain a national forecast. These forecasts could then be compared with the forecasts on the national time series.

Unfortunately, the state (and therefore the regional) data is available only for traffic fatalities. Non-traffic fatalities are estimated on a national basis for inclusion into the national time series for highway fatalities. Thus the summed regions forecast must be compared with a forecast derived from the national traffic fatalities time series.

At first, following the procedures of the previous study, the analyses were performed using 12 years and 6 months of data to obtain a forecast for the last 6 months of calendar year 1972. From the known first 6 months totals and the forecast for the last 6 months forecasts were obtained for the highway traffic fatalities for the entire calendar year 1972.

Midway through the analyses, we were asked to provide forecasts for the calendar year 1973 by each of the several methods under review. This posed no problem as far as the methods using **exponential** smoothing were concerned. However, because the NHTSA's procedures required some data for the first several months of the calendar year which were not then available, the current NHTSA methods required slight modification. The details of these modifications are discussed in section 3.



In the first study<sup>1</sup> the analyses were based on the first 3 months of calendar year data, then the first 6 months of data, and finally the first 9 months of data. In the new work, in order to cut down on the amount of computing required, it was decided to make a comprehensive comparison based on the first 6 months data for a calendar year. We remark that in the previous work, no systematic differences were noted between the results based on 3 and 9 months of data and those based on 6 months of data. Also, rather than carrying out a large number of runs in the time-sharing mode using a typewriter output, we decided to make the comparisons using the GEXSMO time-series analysis routine and have an additional partial comparison using the EXPsmoothing<sup>2</sup> program operated in a time-share mode but with high speed output from tape.

The notations adopted for the previous report will be used in this report; any changes or additions will be noted explicitly. (See List of Symbols and Formulas.)

---

<sup>1</sup>U.S. Department of Commerce, National Bureau of Standards, Time Series Forecasting of Highway Accident Fatalities. NBSIR-73-138 (1973).

<sup>2</sup>For discussion of these time series analysis programs see the discussion and references given in NBSIR-73-138 and the Appendix.

## 2.0 ANALYSES BASED ON 12 YEARS AND 6 MONTHS OF DATA (12, 6)

Production runs were made using monthly (M) data, quarterly (Q) data, half-year (H) data and yearly (Y) data for several levels of aggregation and for two different modes of aggregation.

The first mode consists of making individual analyses and predictions for each state, then sum the results over the states constituting a region, then sum the results of each of the regions in order to obtain a forecast for the national values. This mode is denoted by SS(M), SS(Q), ....

The second mode of aggregation is called lumping and is denoted by L(M), L(Q), .... In this mode, a time series for each region was created by lumping (adding) the data from each state in the region. Forecasts were then obtained for each of the regions directly. Finally, forecasts were obtained for the national values by summing the results of the regions. This latter is denoted by  $\sum$  Regions.

In the summary tables GEXS is used to denote the GEXSMO program, EXPS, the EXPSSMOOTHING program. N denotes an NHTSA method. For each region (U.S.) the forecast for the current calendar year from 6 months of data by a method N is independent of how the aggregation for the region (U.S.) is made. However, the forecasts for region (U.S.) by summing forecasts of states by a method N does not necessarily equal the forecasts for region (U.S.) using lumped data by method N. Thus, for a fixed K,  $N_1^{(K)} \neq N_2^{(K)}$ , where K denotes the number of years of past data used in the N method, and the subscripts refer to the lumping or Sum of States mode, respectively. (See Appendix A for more detail.)

Table 1 contains a summary of the comparisons of forecasts by data features based on the fractional error criterion,  $\Delta/S^*$ . Here  $\Delta_6$  is the error in the forecast for the last 6 months of the calendar year and  $S^*$  denotes the observed value of the fatalities for the calendar year. The forecasts in Table 1 are for the calendar year 1972, based on 12 years, 6 months of data (12, 6). The (signed) fractional error for each region is shown for each mode and method of aggregation.

For the forecasts produced by the GEXSMO program:

- 1) the best forecasts for regions is by L(Q);
- 2) the best forecasts for U.S. is for  $\sum$  Regions by L(Q);
- 3)  $\sum$  Regions forecasts are better than the forecasts for the entire U.S. (based on one grand lumping).
- 4) In general, for regions the mean absolute fractional error (MAE) for lumped forecasts is less than the MAE for forecasts by Sum of States mode. The exception is SS(M).
- 5) In general, for  $\sum$  Regions, the error for forecasts by the lumping mode is less than that for forecasts by Sum of States mode. The exception is SS(M).

Guided by these results we decided to make some computer runs using a different computer time-series analysis program, namely EXPSMOOTHING. Rather than run a complete duplication of the GEXSMO runs, it was decided to operate in the lumped (L) mode only and to use only quarterly, half year and yearly data. The results of these computations are shown in Table 1 also.

Table 1. Comparisons of Estimates by Data Features Based On  $\Delta_6/S^*$  (Fractional Error) Estimates For Calendar 1972, Based on 12 Years, Six Months Data.

Region	L(M)			SS(M)			L(Q)			SS(Q)		
	GEXS	$N_1^{(12)}$	EXPS	GEXS	$N_2^{(12)}$	EXPS	GEXS	$N_1^{(12)}$	EXPS	GEXS	$N_2^{(12)}$	EXPS
1	.0070	-.0023	-	.0135	.0019	-	.0140	-.0023	.0131	.0196	.0019	-
2	.0202	.0091	-	.0191	.0091	-	.0175	.0091	.0191	.0140	.0091	-
3	.0102	.0135	-	-.0081	.0145	-	-.0010	.0135	-.0131	-.0044	.0145	-
4	.0238	.0256	-	.0134	-.0249	-	.0218	.0256	.0161	.0143	-.0249	-
5	.0134	.0370	-	.0223	.0663	-	.0178	.0370	.0304	.0262	.0663	-
6	.0159	.0291	-	.0184	.0277	-	.0208	.0291	.0172	.0159	.0277	-
7	-.0259	.0198	-	-.0242	-.0204	-	-.0198	.0198	.0009	-.0265	-.0204	-
8	-.0220	.0202	-	-.0325	.0188	-	-.0197	.0202	-.0463	-.0247	.0188	-
9	.0037	.0168	-	.0079	.0145	-	.0047	.0168	.0071	.0103	.0145	-
10	.0126	.0010	-	-.0101	.0040	-	.0020	.0010	-.0045	-.0131	.0040	-
Low AE	.0037	.0010	-	.0079	.0019	-	.0010	.0010	.0009	.0103	.0019	-
Mean AE	.0160	.0223	-	.0157	.0280	-	.0154	.0223	.0168	.0169	.0280	-
High AE	.0259	.0370	-	.0325	.0663	-	.0218	.0370	.0463	.0265	.0663	-
$\Sigma$ Regions	.0110	.0176	-	.0081	.0220	-	.0012	.0176	.0111	.0096	.0220	-

	GEXSMO (GEXS) (12,6)						(12,0)		
	L(M)	SS(M)	L(Q)	SS(Q)	L(H)	SS(H)	L(Y)	SS(Y)	
Low AE	.0037	.0079	.0010	.0103	.0015	.0116	.0027	.0026	
Mean AE	.0160	.0157	.0154	.0169	.0188	.0195	.0282	.0282	
High AE	.0259	.0325	.0218	.0265	.0293	.0283	.0655	.0756	
$\Sigma$ Regions	.0110	.0081	.0012	.0096	.0163	.0116	.0076	-.0104	
Entire U.S.	.0126	-	.0087	-	.0181	-	-.0315	-	

(1) Note: For each state (region, U.S.) the forecast for current calendar year from 6 months by method N is independent of how the aggregation for region is made.  
(2) Forecast for region (U.S.) by summary forecasts of states by method N does not necessarily equal forecast for region (U.S.) using lumped data from states by method N.

Table 1. (Continued)

Region	L(H)			SS(H)			L(Y) (12,0)			AA(Y) (12,0)		
	GEXS	N <sub>1</sub> <sup>(12)</sup>	EXPS	GEXS	N <sub>2</sub> <sup>(12)</sup>	EXPS	GEXS	N <sub>1</sub> <sup>(12)</sup>	EXPS	GEXS	N <sub>2</sub> <sup>(12)</sup>	EXPS
1	.0219	-.0023	0	.0224	.0019	-	.0238	-.0023	.0076	.0210	.0019	-
2	.0064	.0091	-.0095	.0153	.0091	-	.0532	.0091	.1065	.0610	.0091	-
3	-.0015	.0135	-.0158	-.0133	.0145	-	-.0298	.0135	-.0601	-.0491	.0145	-
4	.0326	.0256	.0049	.0169	-.0249	-	-.0275	.0256	.0163	-.0172	-.0249	-
5	.0260	.0370	.0221	.0283	.0663	-	.0139	.0370	.0621	-.0047	.0663	-
6	.0293	.0291	.0157	.0238	.0277	-	-.0267	.0291	-.0119	-.0026	.0277	-
7	-.0210	.0198	-.0221	-.0239	-.0204	-	-.0187	.0198	.0216	-.0293	.0204	-
8	-.0261	.0202	-.0243	-.0247	.0188	-	-.0655	.0202	-.1343	-.0756	.0188	-
9	.0116	.0168	.0032	.0145	.0145	-	-.0027	.0168	-.0097	-.0129	.0145	-
10	.0116	.0010	.0080	-.0116	.0040	-	.0201	.0010	-.0090	-.0090	.0040	-
Low AE	.0015	.0010	0	.0116	.0019	-	.0027	.0010	.0076	.0026	.0019	-
Mean AE	.0188	.0223	.0126	.0195	.0280	-	.0282	.0223	.0439	.0282	.0280	-
High AE	.0293	.0370	.0243	.0283	.0663	-	.0655	.0370	.1343	.0756	.0663	-
ΣRegions	.0163	.0176	.0026	.0116	.0220	-	-.0076	.0176	.0118	-.0104	.0220	-

EXPONENTIAL SMOOTHING METHODS  
EXPSMOOTHING (EXPS) (12,6)

(12,0)

	(12,6)						(12,0)		
	L(M)	SS(M)	L(Q)	SS(Q)	L(H)	SS(H)	L(Y)	SS(Y)	
Low AE	-	-	.0009	-	0	-	.0076	-	-
Mean AE	-	-	.0168	-	.0136	-	.0439	-	-
High AE	-	-	.0463	-	.0243	-	.1343	-	-
Σ Regions	-	-	.0176	-	.0026	-	.0118	-	-
Entire U.S.	-	-	.0043	-	.0042	-	-.0060	-	-

Of the EXPSMOOTHING methods, the best forecasts were obtained using half-year data, L(H). For this mode, L(H), the  $\sum$  Regions forecast was better than the L(H) forecast for the Entire U.S., although both were very good. For both quarterly and yearly data, however, the Entire U.S. estimates were better than the  $\sum$  Regions estimates. However, the relatively poorer forecasts of  $\sum$  Regions in this case of L(Q) is due primarily to a poor estimate for one region, Region 8. Very poor estimates were obtained were obtained for Regions 2, 3 and 8 using L(Y). We remark that in general Region 8 was poorly forecast by all methods.

Using the NHTSA method and K=12 (see Appendix) the forecasts for  $\left. \begin{array}{l} \sum \text{Regions} \\ \sum \text{Regions} \end{array} \right\}$  by lumping were better than those produced by the Sum of States mode (see Table 2), and the  $\sum$  Regions forecast by lumping was better than that for the Entire U.S.

Applying the NHTSA methods to the Entire U.S., it was possible to produce 12 different forecasts, depending on whether 1, 2, ..., 12 years of past data was employed. These ranged in value from 55224 to 55930 with a mean of **55593 for the U.S. traffic fatalities.** The corresponding fractional absolute errors ranged from 0.0061 to 0.0190 with a mean of 0.0128. The values of K corresponding to

these forecasts were  $K=4$  and  $K=11$  with the  $K$ -value corresponding to the mean between 6 and 7. (Table 2)

The NHTSA methods<sup>1</sup> may be called ratio methods, since they may be written

$$\frac{\hat{Y}_0^{(K)}}{m_0} = \lambda \frac{\sum_{i=1}^k Y_{-i}}{\sum_{i=1}^k m_{-i}}$$

Forecasts were also made by a difference method

$$Y_0 - m_0 = Y_{-1} - m_{-1} \quad \text{or} \quad Y_0 = m_0 + (Y_{-1} - m_{-1})$$

If  $m_0$  is the sum of data for the first 6 months of the current year, then  $Y_{-1} - m_{-1}$  is the sum of the data for the 1st 6 months of the previous year. Results for 1972 by this method were excellent, the fractional absolute error being 0.0005. This appears to be a fortuitous result. When the method was applied to previous years a wide range of values resulted. (see Table 3)

---

<sup>1</sup>George Suzuki points out that the current NHTSA methods utilizes the so-called "non-seasonal" or "non-decompositional" procedure. This is in contrast to time-series analysis which generally goes through a decomposition process to isolate the trend. In this decomposition process, the seasonal and other regular variations are identified. The reconstitution of a future time series, i.e., extrapolation, is performed by extrapolating the trend and then imposing the corresponding cyclical variations. Both the GEXSMO and EXPSMOOTHING time-series programs do the latter. (See Appendix.)

Table 2. Errors of Estimates of 1972 U.S. Traffic Fatalities by Various Non-Decompositional Methods

Using 12 Years, 6 Months of Data

Region	$\hat{Y}_0$	L		SS	
		N (12) 1	N (12) 2	N (12) 1	N (12) 2
1	.0224	-.0023	.0019		
2	.0016	.0091	.0091		
3	-.0174	.0135	.0145		
4	-.0013	.0256	-.0249		
5	.0252	.0370	.0663		
6	.0038	.0291	.0277		
7	-.0173	.0198	-.0204		
8	-.0417	.0202	.0188		
9	-.0136	.0168	.0145		
10	-.0035	.0010	.0040		
Low AE	.0013	.0010	.0019		
MAE	.0148	.0223	.0280		
HAE	.0417	.0370	.0663		
∑ Region	.0005	.0176	.0220		
Entire US	.0005	.0187	--		

$$\hat{Y}_0 = m_0 + (Y_{-1} - m_{-1})$$

$$N_1^{(k)} = Y_0^{(k)} = \lambda m_0$$

$$\frac{\sum_1^k Y_{-i}}{\sum_1^k m_{-i}}$$

Entire U.S. by $\hat{Y}_0^{(k)}$ (1972)			
	L	M	H
Number	55224	55594	55930
$\Delta/S^*$	.0061	.0128	.0190



Table 3. Estimates of U.S. Traffic Fatalities for 1961-1972 Using a Difference Method:

$$Y_0 = m_0 + (Y_{-1} - m_{-1})$$

Year	First 6 Months $m_0$	$Y_{-1} - m_{-1}$	Forecast $\hat{Y}_0$	Observed $Y_0 = S^*$	$\Delta$	$\Delta/S^*$
1961	16341	20466	36807	37012	- 205	-.0055
1962	17401	20671	38072	39612	-1540	-.0389
1963	18493	22211	40704	42472	-1768	-.0416
1964	20583	23979	44562	46426	-1864	-.0388
1965	21166	25483	47009	47914	- 905	-.0189
1966	23004	26748	49752	51616	-1864	-.0361
1967	22963	28612	51575	51648	- 73	-.0014
1968	24319	28685	53004	53799	- 795	-.0148
1969	25022	29480	54502	54953	- 451	-.0082
1970	24145	29931	54076	53207	869	.0163
1971	23922	29062	52984	53768	- 784	-.0146
1972	25073	29846	54919	54889	30	.0005

## 2.2 A Look to Regions

On the basis of the forecasts for 1972 using 12 years and 6 months of data, it appeared that some improvement could be obtained in estimating **national traffic fatalities**, if the estimates for each region could be improved. There also appears to be some legitimate merit in making accurate estimates for the regions in their own right.

Two places to look for possible improvements are (a) those regions which contribute the largest numbers to the national total and (b) those regions which were forecast poorly.

Over 50 percent of the national traffic fatalities for 1972 were contributed by Regions 4, 5 and 6. (See Table 4). Regions 9 and 3 contribute another 20 percent of the total. If one were to look for improvements in forecasting **national traffic fatalities**, these are the regions that should be examined for homogeneity of data characteristics.

If Regions 4, 5 and 6 were forecast accurately, by say the GEXSMO program using the lumped mode and the other Regions remained unchanged, then the total error in the forecast could be reduced from 604 to 70 (out of 54889) if monthly data were used. The reduction in error would be from 478 to -110, from 993 to 234 and from -417 to -101 if quarterly, half-year or yearly data were employed, respectively. These might be looked upon as upper bounds, it being presumptuous to assume that the forecast could be made with no error. However, if the accuracy of the forecasts for these three

Table 4. Order of Region by Number of 1972 Traffic Fatalities

Order	Region	1972 Fatalities	Percent of Total	Cum %
1	4	11998	21.86	21.86
2	5	10363	18.88	40.74
3	6	6862	12.50	53.24
4	9	6189	11.28	64.52
5	3	5177	9.43	73.95
6	2	4508	8.21	82.16
7	7	3477	6.33	88.49
8	8	2182	3.98	92.47
9	1	2143	3.90	96.37
10	10	1990	3.62	99.99
	Total	54889		

regions alone were improved by a factor of 2, i.e. from slightly over 2 percent average error to slightly over 1 percent average error, then the total error in the national forecast would have been reduced to 137, 184, 663, -259 by the four data aggregation M, Q, H and Y, respectively, using the GEXSMO program.

Before any calculations were made, the NHTSA time-series program TIMSR<sup>4</sup> output was examined by state for each region. The seasonality factors were scanned to determine if a particular region could be broken down into sub-regions that had similar seasonality factor patterns. The results of this exercise are given in Table 5. Using these definitions of sub-regions as a guideline, calculations of regional fatalities were made using the GEXSMO program with lumped quarterly, L(Q), data for the sub-regions 4.2, 5.1 and 5.2. Since subregion 4.1 consisted of the one state (Florida), this calculation was at hand in the SS(Q) set of calculations. Adding the results of fatalities for regions 4.1 and 4.2 the new estimate of fatalities for Region 4 became 12126, an overforecast by 83 fatalities with fractional error  $\Delta/S^* = .0178$  in both cases. These are an improvement from  $\Delta = 272$  to  $\Delta = 183$  when comparison was made to the SS(Q) mode for Region 5.

No recalculations were made for Region 6, the region with the third highest fatality total for 1972, because there did not appear to be any sub-regions distinguishable from the entire region on the basis of seasonality factors.

Table 5. Tentative New Regions Based on Similarity of Seasonality Factors from TIMSR4

Region*	New Region	States
4	4.1	Florida
	4.2	Others: Alabama, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee
5	5.1	Minnesota, Wisconsin
	5.2	Illinois, Indiana, Ohio, Michigan
6	6.0	No change; Arkansas, Louisiana, New Mexico, Oklahoma, Texas
9	9.1	California, Arizona
	9.2	Hawaii
	9.3	Nevada
3	3.1	District of Columbia
	3.2	Others: Delaware, Maryland, Pennsylvania, Virginia, West Virginia
2	2.0	No change: New Jersey, New York, Puerto Rico**
7	7.0	No change: Iowa, Kansas, Missouri, Nebraska
8	8.0	No change: Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming
1	1.1	Maine, Vermont
	1.2	Rhode Island
	1.3	Connecticut, Massachusetts, New Hampshire
10	10.1	Alaska
	10.2	Idaho
	10.3	Oregon, Washington

\* In order of decreasing magnitude of traffic fatalities for 1972.

\*\* Puerto Rico not included in this study.

Table 6. Fractional Error of Regional Estimates for 1972 by Various Methods  
Based on 12 Years, 6 Months of Data

GEXS	Region										MAE	Frequency of Best (B)
	1	2	3	4	5	6	7	8	9	10		
L(M)	.0070B	.0202	.0102	.0238	.0134	.0159	-.0259	-.0220	.0037	.0126	.0160	1
SS(M)	.0135	.0191	-.0081	.0134B	.0223	.0184	-.0242	-.0325	.0079	-.0101	.0157	1
L(Q)	.0140	.0175	-.0010B	.0218	.0178	.0208	-.0198	-.0197B	.0041	.0020B	.0154	3
SS(Q)	.0196	.0140	-.0044	.0143	.0262	.0159	-.0265	-.0247	.0103	-.0131	.0169	0
L(H)	.0219	.0064B	-.0015	.0326	.0260	.0393	-.0210	-.0261	.0116	.0116	.0188	1
SS(H)	.0224	.0153	-.0133	.0169	.0283	.0238	-.0239	-.0247	.0145	-.0116	.0195	0
L(Y)	.0238	.0532	-.0299	-.0275	.0139	.0267	-.0187B	-.0655	-.0027B	.0201	.0282	2
SS(Y)	.0210	.0610	-.0491	-.0172	-.0047B	-.0026B	-.0293	-.0756	-.0125	-.0090	.0282	2
MAE8	.0179	.0258	.0147	.0209	.0191	.0192	.0230	.0364	.0085	.0113	.0198	
Rank	4	9	3	7	5	6	8	10	1	2		
EXPS												
L(M)	-	-	-	-	-	-	-	-	-	-	-	
SS(M)	-	-	-	-	-	-	-	-	-	-	-	
L(Q)	.0131	.0191	-.0130B	.0161	.0304	.0172	.0009B	-.0463	.0071	-.0045B	.0168	3
SS(Q)	-	-	-	-	-	-	-	-	-	-	-	
L(H)	0	B	-.0095B	.0158	.0049B	.0221B	.0157	-.0221	-.0243B	.0032B	.0126	6
SS(H)	-	-	-	-	-	-	-	-	-	-	-	
L(Y)	0076	.1065	-.0601	.0163	.0621	-.0119B	.0216	-.1343	-.0097	-.0090	.0439	1
SS(Y)	-	-	-	-	-	-	-	-	-	-	-	
MAE3	.0069	.0450	.0297	.0124	.0382	.0149	.0149	.0683	.0067	.0072	.0244	
Rank	2	9	7	6	8	4T	4T	10	1	3		
N												
N <sub>1</sub> (12)	-.0023	.0091B	.0135B	.0256	.0370B	.0291	.0198B	.0202	.0168	.0010B	.0223	5
N <sub>2</sub> (12)	.0019B	.0091B	.0145	-.0249B	.0663	.0277B	-.0204	.0188B	.0145B	.0040	.0280	6
MAE2	.0021	.0091	.0140	.0252	.0516	.0284	.0201	.0195	.0156	.0020	.0252	
Rank	2	3	4	8	10	9	7	6	5	1		

Note: T=Tie; B=Best forecast for region by methods under consideration

Region 9, the fourth highest in 1972, was the best forecast region by both the GEXSMO and EXPONENTIAL programs. There did not appear to be any great chance of picking up much improvement in forecast in this region.

Table 6 gives a comparison of forecasts for 1972 by region by each of the methods using 12 years and 6 months of data. The comparison are on the basis of the fractional error ( $\Delta/S^*$ ). In this table, which contains the data of Tables 1 and 2 but is presented to emphasize regions, entries are marked B to indicate the "best" mode of forecasting each regions fatalities. "Best" mode is that for which the absolute value of the fraction error is smallest. On this basis, for the GEXSMO program forecasts, the best mode of forecasting is L(M) for Region 1, L(H) for Region 2, L(Q) for Region 3, SS(M) for Region 4, etc. The most frequent "best" mode is L(Q) with a frequency of 3 out of 10. The mode with the lowest mean absolute fractional error is also L(Q). The modes with no best regional forecasts are SS(Q) and SS(H). Region 8 had the worst "best" forecast with  $\Delta/S^*$  for this region being  $-.0197$ . On the basis of mean absolute fractional error the region that had best forecasts is Region 9.

The ranking order for each of the regions is also given in Table 14. The symbol, 4 , means that (in this case Region 1) was the 4th best forecast region; 10 indicates the worst forecast on the average. This dubious distinction went to Region 8 for all three methods: GEXSMO, EXPSMOOTHING and NHTSA methods.

While 8 modes of forecasting were done by GEXSMO, only three modes were run using the EXPSMOOTHING program: L(Q), L(H) and L(Y). On the basis of frequency of best regional forecast, L(H) had a frequency of 6, L(Q) had 3 and L(Y) had 1.

Results for the two possible NHTSA methods, by sums of states method ( $N_2$ ) and the lumping method ( $N_1$ ) were fairly evenly divided: 6 to 5 in favor of  $N_2$ .

The frequency of firsts, seconds and third "best" regional forecasts by procedure is given in Table 7

TABLE 7

Frequency of First, Second and Third Best Procedure (Regions)

	GEXS	EXPS	N
First	5	3	2
Second	4	4	2
Third	1	3	6

On the basis of this sample, if one were interested in accuracy of regional forecasts, it would appear that one of the time series programs should be preferred over the current N methods.



The second way to improve forecasts would be to improve the accuracy of the worst forecast regions, provided of course that the improvement would be large enough to seriously affect the national values. Since the current national traffic fatalities is near 55,000, a change of 550 would affect a change of approximately one percent in the national forecast, 275 would affect a change of approximately 0.5 percent and 110 would effect a change of approximately 0.2 percent. In our notation the fractional error would be .01, .005 and .002 respectively.

Region 8 was the worst forecast region by each of the procedures. The total fatalities for 1972 was 2182 or approximately 4 percent of the national value. The best forecast by any method is close to 2 percent or  $\approx$  44 fatalities. But this is less than 0.1 of a percent of the national value, so we abandoned any effort to make special progress on this region. We remark in passing that Region 8 was underforecast by all of the GEXSMO and EXPSMOOTHONG programs, and that Region 7 was underforecast by all the GEXSMO programs.

The second worst forecast region is Region 2 which contributes about 8 percent of the national total. However, this region was forecast very well by GEXSMO using L(H) and by EXPSMOOTHING using L(H) and by both N methods. Table 5 indicates there is no need to break the region into subregions. Also this case has already been studied by the sum of states mode since the region consists of only 2 states:<sup>1</sup> New York and New Jersey.

---

<sup>1</sup>Puerto Rico was not included in this study.

### 3.0 FORECASTS OF TOTAL U.S. TRAFFIC FATALITIES USING NO DATA FROM CURRENT YEAR

#### 3.1 Modifications Needed

To provide forecasts for the calendar year 1973, two things were necessary: (1) the need to modify the NHTSA method slightly since data for the current calendar year are not available and (2) to obtain some forecasts for previous years using these methods and forecasts by the exponential smoothing methods using no current year data. The latter is needed to get some feel for the error involved the 1973 forecasts.

To obtain forecasts for this year based on last years data two modifications of the NHTSA method were used. The first was

$$\frac{\hat{Y}_{0+1}^{(K)}}{m_0} = \frac{\frac{1}{K} \sum_{i=1}^K Y_{-i+1}}{\frac{1}{K} \sum_{i=1}^K m_{-i}} = \frac{\sum_{i=1}^K Y_{-i+1}}{\sum_{i=1}^K m_{-i}} \quad K = 1, \dots$$

The right side of this equation represents the ratio of average pairs.

For 1973,  $K = 12$ , while for 1972,  $K = 11$ . The second method was

$$\frac{\hat{Y}_{0+1}^{(K)}}{m_0} = \frac{1}{K} \sum_{i=1}^K \frac{Y_{-i+1}}{m_{-i}} = \text{average of } K \text{ ratios, } K = 1, \dots$$

Each of these methods generates  $K$  estimates

$$\hat{Y}_{0+1}^{(1)}, \hat{Y}_{0+1}^{(2)}, \dots, \hat{Y}_{0+1}^{(K)}$$

and one can take the mean of these  $K$  estimates to provide other estimates.

Using difference techniques we obtained

$$\hat{Y}_{0+1} - m_0 = Y_0 - m_{-1} \quad \text{or} \quad \hat{Y}_{0+1} = Y_0 + (m_0 - m_{-1})$$

A simple variant on this is

$$\hat{Y}_{0+1}^{(K)} - m_0 = \frac{1}{K} \sum_{i=0}^{K-1} (Y_i - m_{-i+1}) = \overline{Y_0 - m_{-1}} \quad K = 1, \dots$$

Another is

$$\hat{Y}_{0+1}^{(K)} = Y_0 + \frac{1}{K} \sum_{i=0}^{K-1} (m_{-i} - m_{-i+1}) = Y_0 + \overline{m_0 - m_{-1}}, \quad K = 1, \dots$$

Results of employing some of these modifications to obtain forecasts for the current year and previous years are presented in Tables 8-12.

In each of these tables  $m$  refers to the lumped data for the last 6 months of the year and  $Y$  refers to the yearly total.

Forecasts for 1973, 1972, and 1971 using the method of the ratio of  $K$  average pairs are presented in Tables 8 and 9. For 1973, the low value is obtained for  $K = 3$ , the high value for  $K = 11$  and the  $K$  corresponding to the mean value is between 6 and 7. For 1972, the low value is obtained for  $K = 2$ , the high for  $K = 10$  and the  $K$  corresponding to the mean value is between 5 and 6. For 1971, the values are  $K = 1$ ,  $K = 9$ , and between 7 and 8. For both 1971 and 1972 the observed value is located in the same interval as the mean value of the  $K$  forecasts.

Similar results were obtained using the method of the average of  $K$  ratios. See Tables 10 and 11.

Results by the simple difference method (modified) for the years 1962-1973 are presented in Table 12.

To reduce the amount of computation the exponential smoothing programs were run only in the lumped mode. Forecasts for 1973 are presented in Table 13, while corresponding forecasts and errors for 1972 are given in Tables 14 and 15. In addition to forecasting for

Table 8. Computation of Forecasts of U.S. Traffic Fatalities for 1973 by a

Modified NHTSA Method:  $\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{\sum_{i=1}^k Y_{-i+1}}{\sum_{i=1}^{m_0} Y_{-i}} = \text{ratio of } k \text{ average pairs}$

Year	Last 6 Months $m_0$	$\sum_{i=1}^{m_0} Y_{-i}$ (2)	Y=S*	$Y_{0+1}$	$\sum_{i=1}^k Y_{-i+1}$ (5)	$(5) \div (2)$	k	1973 $\hat{Y}_{0+1}^{(k)}$
1960	20466	315534	37022	37012	587316	1.86134	12	55498
1961	20671	295068	37012	39612	550304	1.86501	11	55607H
1962	22211	274397	39612	42472	510692	1.86114	10	55492
1963	23979	252186	42472	46426	468220	1.85664	9	55358
1964	25843	228207	46426	47914	421794	1.84830	8	55109
1965	26748	202304	47914	51616	373880	1.84811	7	55103M
1966	28612	175616	51616	51648	322264	1.83505	6	54714
1967	28685	147004	51648	53799	270616	1.84088	5	54888
1968	29480	118319	53799	54953	216817	1.83248	4	54638
1969	29931	88839	54953	53207	161864	1.82199	3	54324L
1970	29062	58908	53207	53768	108657	1.84452	2	54996
1971	29846	29846	53768	54889	54889	1.83908	1	54834
1972	29816		54889					
							$\sum$	660561
							Low	54324
							Mean	55048
							High	55607

Table 9. Computation of Estimates of U.S. Traffic Fatalities for 1971, 1972

by a Modified NHTSA Method:  $\frac{\hat{Y}_{0+1}^{(k)}}{m_0} = \frac{\sum_1^k Y_{-i+1}}{\sum_1^k m_{-i}}$  = ratio of k average pairs

Year	1972				1971				
	$\sum_1^k m$	$\sum_1^k Y_{0+1}$	$\frac{\hat{Y}_{0+1}}{m_0}$	k	$\hat{Y}_{0+1}^{(k)}$	$\sum_1^k Y_{0+1}$	$\frac{Y_{0+1}}{m_0}$	k	$\hat{Y}_{0+1}^{(k)}$
1960	285688	532427	1.86366	11	55623	478669	1.86524	10	54208
1961	265222	495415	1.86792	10	55750 ← H	441647	1.87012	9	54349 ← H
1962	244551	455803	1.86384	9	55628	402035	1.86569	8	54221
1963	222340	413331	1.85900	8	55484	359563	1.86034	7	54065 ← M, S*
1964	198361	366905	1.84968	7	55206	313137	1.84961	6	53753
1965	172518	318991	1.84903	6	55186 ← M, S*	265223	1.84881	5	53730
1966	145770	267375	1.83422	5	54744	213607	1.83027	4	53191
1967	117158	215727	1.84133	4	54956	161959	1.83843	3	53428
1968	88473	161928	1.83025	3	54626	108160	1.82053	2	52908
1969	58993	106975	1.81335	2	54121 ← L	53207	1.77766	1	51662 ← L
1970	29062	53768	1.85011	1	55218				
1971	29846								
				$\sum$	606542				535515
				Low	54121				51662
				Mean	55140				53552
				High	55750				54349
				S*	54889				53768
				$\Delta L$	- 768				-2106
				$\Delta M$	251				216
				$\Delta H$	861				540
				$\Delta L/S^*$	-.0140				-.0392
				$\Delta M/S^*$	.0046				.0040
				$\Delta H/S^*$	.0157				.0098





Table 12. Computation of Estimates of U. S. Traffic Fatalities for 1962-1973 Using a Modified

Difference Method:  $\hat{Y}_{0+1} - m_0 Y_0 - m_0^{-1} = Y_{0+1} = Y_0 + (m_0^{-1} - 1) Y_0$

Year	Last 6 months $m_0$	$m_0^{-1}$	$Y_0$	$\hat{Y}$	$\hat{Y} - Y_0$	$\Delta/Y_0$
1960	20446	-	37022	-	-	-
1961	20671	+ 225	37012	-	-	-
1962	22211	+1540	39612	37237	-2375	-.0600H
1963	23979	+1768	42472	41152	-1320	-.0311
1964	25843	+1864	46426	44240	-2188	-.0471
1965	26748	+ 905	47914	48290	+ 376	.0078
1966	28612	+1864	51616	48819	-2797	-.0542
1967	28685	+ 73	51648	53480	+1832	-.0355
1968	29480	+ 795	53799	51721	-2078	-.0386
1969	29931	+ 433	54953	54594	+ 359	.0065
1970	29062	- 869	53207	53386	=2129	-.0400
1971	29846	+ 784	53768	52338	-1430	-.0266
1972	29816	- 30	54889	54552	- 337	-.0061L
1973			?	54859		



Table 13. Forecasts of U.S. Traffic Fatalities for 1973  
Using a Variety of Methods

Region	GEXSMO				EXPSMOOTHING					
	L(M)	L(Q)	L(H)	L(Y)	L*	L(M)	L(Q)	L(H)	L(Y)	L <sub>1</sub> *
Entire U.S.	54660	54186	54698	54821	54591	-	55994	55551	55392	55646
Σ Regions	54808	54569	54888	55489	54938	-	55797	55537	54643	55325
1	2180	2186	2196	2178	2185	-	2227	2227	1973	2142
2	4691	4679	4649	4745	4691	-	4746	4640	4638	4675
3	5349	5345	5215	5214	5281	-	5276	5187	5089	5184
4	12080	11785	11980	12327	12043	-	12160	12186	12077	12141
5	9845	9842	10057	10118	9966	-	10462	10200	10167	10276
6	6693	6789	6840	6982	6826	-	6883	6967	6927	6926
7	3548	3547	3518	3459	3518	-	3571	3660	3416	3549
8	2122	2154	2148	2091	2129	-	2145	2157	2135	2146
9	6220	6242	6255	6304	6255	-	6276	6263	6144	6244
10	2080	2000	2030	2071	2045	-	2051	2050	2027	2043

NHTSA (Modified): Entire U.S.

Ratio of K average pairs      Average of K ratios      Difference (Modified)

27      L      M      H      L      M      H

54328      55048      55607      54333      55084      55578      54859

Low of 25 Estimates      54186, Σ Regions, L(Q), GEXSMO

Grand Average of 25 Estimates      55055

High of 25 Estimates      55994, Entire U.S., L(Q), EXPSMOOTHING

\*  $\bar{L} = \frac{1}{4} (L(M)+L(Q)+L(H)+L(Y))$

$\bar{L}_1 = 1/3 (L(Q)+L(H)+L(Y))$

each region,  $\Sigma$  Regions and Entire U.S. by M, Q, H, and Y an average of these four methods was taken. This is denoted by  $\bar{L}$ , where

$$\bar{L} = \frac{1}{4}[L(M) + L(Q) + L(H) + L(Y)].$$

Using GEXSMO, the best estimate for the national traffic fatalities for 1972 was obtained using L(H), and identical estimates were obtained for  $\Sigma$  Regions and Entire U.S. by L(H). However, all of the forecasts for 1972 by  $\Sigma$  Regions and Entire U.S. had percent error less than 1 percent, with the exception being Entire U.S. by L(Y). Excluding this outlier of  $-.0315$ , all the GEXSMO estimates were in the interval  $[.0019 \text{ to } .0089]$ .

Using EXPSMOOTHING the best estimate for the national traffic fatalities for 1972 was obtained using  $\Sigma$  Regions and L(Q). (Tables 14, 15)

Results for the NHTSA (Modified) method applied to the Entire U.S. for 1972 range from 54121 to 55870 with mean average fractional errors of  $.0046$  and  $.0053$ . (Tables 14, 15)

A grand average of the 25 estimates for 1972 not all independent, slightly underforecast the national traffic fatality value by  $-265$  ( $-.0048$ ). (Tables 14, 15).

Table 14. Estimates of U. S. Traffic Fatalities for 1972 Using 12 Years of Data.

Region	GEXSMO				Actual S*	EXPSMOOTHING				
	L(M)	L(Q)	L(H)	L(Y)		L <sub>1</sub>	L(M)	L(Q)	L(H)	L(Y)
Entire U.S.	54710	55379	54787	53166	54889	54510	53291	53535	54562	53795
Σ Regions	54561	55141	54786	54471	54889	54740	55084	53941	55537	54853
1	2189	2179	2222	2194	2143	2196	2190	2235	2160	2195
2	4717	4710	4673	4748	4508	4712	4643	4647	4988	4759
3	5279	5246	5089	5022	5177	5159	5211	4974	4866	5017
4	11935	12121	11993	11668	11998	11929	11803	11786	12194	11928
5	10298	10668	10427	10507	10363	10475	10757	10368	11007	10711
6	6675	6732	6811	6679	6862	6724	6845	6852	6780	6826
7	3375	3424	3431	3412	3477	3410	3530	3590	3552	3557
8	2058	2037	2054	2039	2182	2047	2019	2064	1889	1991
9	5998	6061	6053	6172	6189	6071	6091	5513	6129	5911
10	2037	1963	2033	2030	1990	2016	1995	1912	1972	1960

NHTSA (Modified): Entire U.S.

Ratio of K Average Paris		Average of K ratios		Difference (Modified)
L	M	L	M	
54121	55140	54137	55179	54552

Low of 25 estimates 53166      Entire U.S., L(Y), GEXSMO  
 Grand Average of 25 estimates 54624  
 High of 25 estimates 55537      Σ Regions, L(Y), EXPSMOOTHING

Table 15. Fractional Errors of Regional Estimates for 1972 by Various Methods

Method	Region										MAE	Σ Regions	Entire U.S.	Freq. of Best	
	1	2	3	4	5	6	7	8	9	10					
GEXS															
L(M)	.0215	.0464	.0197	-.0053	-.0063	-.0273	-.0293	-.0568B	-.0309	.0236	.0267	-.0059	-.0033	1	
L(Q)	.0168B	.0448	.0133B	.0103	.0295	-.0190	-.0152	-.0665	-.0207	-.0136B	.0250	.0046	.0089	3	
L(H)	.0369	.0366B	-.0170	-.0004B	.0062B	-.0074B	-.0132B	-.0587	-.0220	.0216	.0220	-.0019B	-.0019B	5	
L(Y)	.0238	.0532	-.0299	-.0275	.0139	-.0267	-.0187	-.0655	-.0027B	.0201	.0282	-.0066	-.0315	1	
L															
L <sub>1</sub>	.0247	.0453	-.0035	-.0058	.0108	-.0210	-.0193	-.0619	-.0191	.0131	-.0069	-.0027	-.0069		
MAE	.0247	.0453	.0200	.0109	.0140	.0218	.0193	.0619	.0191	.0197	.0048	.0048	.0114		
Rank	8	9	6	1	2	7	4	10	3	5					
EXPS															
L(M)															
L(Q)	-.0219	.0299B	.0066B	-.0163B	.0380	-.0025	.0152B	-.0747	-.0151	.0025B	-.0223	-.0036B	-.0291	4	
L(H)	.0429	.0308	-.0392	-.0176	.0005B	-.0014B	.0325	-.0541B	-.1092	-.0392	.0367	-.0172	-.0249	3	
L(Y)	.0076B	.1065	-.0601	.0163B	.0621	-.0119	.0216	-.1343	-.0097B	-.0090	.0449	.0150	-.0060B	2	
L															
L <sub>1</sub>	.0241	.0557	-.0309	.0059	.0335	-.0053	.0231	-.0877	-.0449	-.0152	.0005	-.0005	-.0200		
MAE	.0241	.0557	.0353	.0167	.0335	.0053	.0231	.0871	.0449	.0136					
Rank	5	9	7	3	6	1	4	10	8	2					

NHTSA (Modified): Entire U.S.

Ratio of K Average Pairs	Average of K Ratios			Difference Modified
	L	M	H	
L	-.0140	.0046	.0157	-.0061
M		-.0137	.0053	
H			.0179	

Low of 25 estimates    -.0315    Entire U.S., L(Y), GEXSMO  
 Grand Average of 25 estimates    -.0048  
 High of 25 estimates    .0150    Σ Region, L(Y), EXPSMOOTHING

Note: B = Best Forecast of L(M), L(Q), L(H), L(Y)

#### 4.0 CONCLUSIONS

Based on the computations using 12 years and 6 months of data to estimate the national highway traffic fatalities for the last 6 months of 1972 (and thus the total calendar year traffic fatalities), there appears to be some advantage in using lumped data for each region aggregated either quarterly or half yearly, and using the Sum of Regions Method to estimate the national value.

There does not appear to be any great difference in the results obtained by using the non-decompositional methods and those obtained by time-series analysis programs based on exponential smoothing methods. Only a few of the vast number of variations possible by non-decompositional methods were tried, and the results of these compared favorably with those done using exponential smoothing. Of the K forecasts possible with the NHTSA methods (or slight modifications) good results were obtained for national estimates by averaging over the K forecasts.

There does not appear to be any great gain to be had in creating new sub-regions of existing regions on the basis of homogeneity of seasonality factors, although some benefit was evidenced. Although not investigated here, possibly a redefinition of the regions by homogeneity of data properties might lead to better forecasts for the nation.

Several regions (notably Regions 8 and 2) were consistently forecast poorly by all methods employed. However, the forecasts for Region 2 were in error by less than one percent by the NHTSA methods  $N_1^{(12)}$  and  $N_2^{(12)}$ . By contrast, Region 9 was forecast well by all methods tried.

On the basis of the sample taken, if one were interested in accuracy of regional forecasts, it would appear that one of the time-series programs should be preferred over the current N methods.

To forecast national highway traffic fatalities for 1973 for non-decompositional methods used by NHTSA had to be modified slightly since no data for the early months of 1973 were at hand. Using these modified methods and the two exponential smoothing methods programmed to yield yearly rather than half-year forecasts, 25 estimates, not necessarily independent, were made for calendar year 1973. These ranged from a low of 54186 to a high of 55994, with mean of 55055. The corresponding estimates by the same methods for 1972 ranged from a low of 53166 to a high of 55537 with a mean of 54624. Observed national traffic fatalities for 1972 were 54889.

APPENDIX  
COMMENTS ON FORECASTING TECHNIQUES AND COMPUTER PROGRAMS

Time series analysis generally goes through a decomposition process to isolate the trend. In this process, the seasonal and other regular variations are identified. The reconstitution of a future time series, i.e., extrapolation, is performed by extrapolating the trend and then imposing the corresponding cyclical variations. Since the NHTSA need is for the annual total, the requirement is less severe since the annual total can be projected without direct consideration of seasonal or other cyclical variations.

The current NHTSA procedures utilizes the "non-seasonal" or "non-decompositional" procedure, but does this in a very straightforward way with one assumption and variants of the assumption. The assumption NHTSA uses is one of proportionality, that is, that this year's fatalities is proportional to last years' fatalities as the cumulative total up to this time is to the corresponding cumulative total last year. The variants have to do with using the average of the last k years of cumulative totals instead of just last year's. Further variations not now among the NHTSA methods can be tried on this, such as various weighted averaging (experimental smoothing) of the last k years just like the weighted schemes in time series analysis.

A different basic assumption would be one of additivity instead of proportionality. The simplest procedure under this assumption is to take the fatalities over the latest one year period. In your notation,  $\hat{Y}_0 = m_0 + (Y_{-1} - m_{-1})$ , where the second term is the fatalities last year over the period complementary to  $m_0$ . Many variations can be

imposed on this, such as various weighting schemes on the prior years' data. This additive scheme is more conservative, but may be most suitable when the third quarter data are in. For the first quarter data, this procedure may be too conservative when there is a significant trend.

Clearly, it is possible to take combinations of the additive and proportional schemes. All of these procedures would be very simple methods.

A further possibility is to combine either of the "non-decompositional" approach with that of time series analysis, principally the trend analysis portion. If the trend projection is imposed on the procedures with the additive assumption the result will be somewhat kin to the procedures using time series analysis. If imposed on the proportionality procedure, the concept will be quite different as might be the results.

Then there is the wide gamut of procedures that utilize other data in addition to the fatalities. Thus, related data such as auto registration, number of teenaged drivers, etc., that might be used as additional predictors would be utilized. This procedure could soon extend outside of the "simple" procedures, but would certainly be the appropriate approach to investigate if the extrapolation extends much beyond the current annual total.

These (edited) comments have been graciously furnished by Dr. George Suzuki of the Technical Analysis Division, National Bureau of Standards. As indicated in the main report and the list of symbols and formulae, several of these methods have been tried, but an exhaustive study was not made.



U.S. DEPT. OF COMMERCE BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBSIR 73-189	2. Gov't Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE  A Comparison of Several Methods for Forecasting U.S. Traffic Fatalities		5. Publication Date	
		6. Performing Organization Code	
7. AUTHOR(S) Alexander R. Crow		8. Performing Organization NBSIR 73-189	
9. PERFORMING ORGANIZATION NAME AND ADDRESS  NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, D.C. 20234		10. Project/Task/Work Unit No. 4314496	
		11. Contract/Grant No.	
12. Sponsoring Organization Name and Address National Highway Traffic Safety Administration Department of Transportation Washington, D.C. 20590		13. Type of Report & Period Covered Final Report	
		14. Sponsoring Agency Code	
15. SUPPLEMENTARY NOTES			
<p>16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)</p> <p>This is a second report to the Mathematical Analysis Division of the National Highway Traffic Safety Administration (NHTSA) on the subject of forecasting annual highway fatalities. This report concerns a comparison of several time series analysis programs based on exponential smoothing and non-decompositional methods currently employed by NHTSA for projecting the annual <u>traffic</u> fatalities for the entire U.S. Several methods of data aggregation are studied.</p> <p>It is found that there is some advantage in using lumped (pooled) data for each region aggregated either quarterly or half yearly, and using the Sum of Regional estimates to estimate the national value.</p> <p>Also, there does not appear to be any great difference in the results obtained using the non-decompositional methods and those obtained by time-series analysis programs based on exponential smoothing methods.</p> <p>Estimates for the 1972 and 1973 national traffic fatalities by a variety of methods were made. For 1973 the estimates ranged from a low of 54186 to a high of 55994, with a mean of 55055.</p>			
17. KEY WORDS (Alphabetical order, separated by semicolons)  Exponential smoothing; forecasting; time-series; traffic fatalities.			
18. AVAILABILITY STATEMENT  <input checked="" type="checkbox"/> UNLIMITED.  <input type="checkbox"/> FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NTIS.		19. SECURITY CLASS (THIS REPORT)  UNCLASSIFIED	21. NO. OF PAGES
		20. SECURITY CLASS (THIS PAGE)  UNCLASSIFIED	22. Price





