NHTSA-84-1

REFERENCE USE ONLY

DOT-HS-806-525
DOT-TSC-NHTSA-84-1

Socio-Economic Influences on Highway Fatalities: An Empirical Investigation

Paul Hoxie
David Skinner
George H. Wang

Transportation Systems Center
Cambridge MA 02142

February 1984
Final Report

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
National Highway Traffic Safety Administration

Office of Research and Development
National Center for Statistics and Analysis
Washington DC 20590

/ cord

Technical Report Documentation Page

1. Report No. DOT-HS-806-525	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle SOCIO-ECONOMIC INFLUENCES ON HIGHWAY FATALITIES: AN EMPIRICAL INVESTIGATION		5. Report Date
		6. Performing Organization Code DTS-45
7. Author(s) Paul Hoxie, David Skinner, and George H. Wang		8. Performing Organization Report No. DOT-TSC-NHTSA-84-1
9. Performing Organization Name and Address U.S. Department of Transportation Research and Special Programs Administration Transportation Systems Center Cambridge MA 02142		10. Work Unit No. (TRAIS) HS470/R4419
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address U.S. Department of Transportation National Highway Traffic Safety Administration Office of Research and Development Washington DC 20590		13. Type of Report and Period Covered FINAL REPORT
		14. Sponsoring Agency Code NRD-31

15. Supplementary Notes

16. Abstract

This study identifies socio-economic variables which are strongly associated with highway fatalities. Further analysis of the relationship between these variables and fatalities reveals that two of the variables, retail sales and personal income, influence fatalities most strongly, meet tests of "statistical causation", and influence fatalities by causing changes in the amount of driving.

In spite of the seemingly anomalous fact that fatalities decreased by ten persent in 1982 while VMT increased by about one percent, models of highway fatalities based on VMT or its proxies, personal income or retail sales, predict fatality declines of 6 to 8.5 percent in 1982. The study analyzes the model for VMT to better understand the complex relationship between VMT and fatalities.

17. Key Words	18. Distribution Statement DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	108	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Preface

This study identifies the socio-economic influences on highway fatalities and investigates the effects of these influences in 1982.

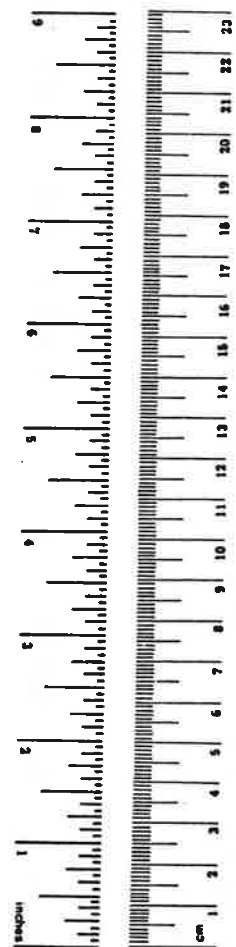
The work was performed by the U.S. Department of Transportation, Research and Special Programs Administration, Transportation Systems Center, Cambrdige, Massachussets, under the sponsorship of the U.S. Department of Transportation, National Highway Traffic Safety Administration, Office of Research and Development, Washington, DC.

The authors are grateful to James Hedlund, Chief of the Mathematical Analysis Division of NHTSA's National Center for Statistics and Analysis, for suggesting and sponsoring this study and for giving many valuable suggestions to improve it. The authors are also indebted to Carol Gurvitz, Peter Mengert, Simon Prensky and Donald Sussman of TSC who made helpful suggestions in their reviews of the work, and to Susan Dresley of the Raytheon Service Company who obtained much of the information, references, and data used by the authors in the study. The authors are also indebted to Carol Arlington and Robin Barnes who provided the indispensable typing support which made the many drafts and revisions possible.

METRIC CONVERSION FACTORS

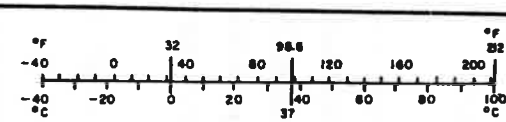
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



EXECUTIVE SUMMARY

Background

In 1982, highway fatalities dropped by more than ten percent from the 1981 level, despite the fact that the estimated total vehicle miles traveled (VMT) increased by slightly more than one percent. This counter intuitive trend began in 1981, when fatalities dropped by more than three percent and VMT increased by almost three percent. Since the economy was in a deep recession during 1981 and 1982, with unemployment rising above ten percent, it has been suggested that economic factors may have caused the fatality decrease.

The purpose of this study is to identify the socio-economic influences on highway fatalities and specifically to investigate the effect of these influences in 1982. The relationships between fatalities and socio-economic variables are explored for the years 1975 through 1982 to see if the 1982 decline is consistent with these longer-term relationships. Many factors influence fatalities which are not purely "socio-economic": restraint and alcohol use, emergency medical services, and automobile and roadway design are among them. Therefore, the analysis performed in this study explains only part of the observed changes in fatalities.

Methodology

Statistical time-series methods are used to analyze the highway fatality data between 1975 and 1982. The objective, identifying socio-economic influences on fatalities, is complicated by the high correlation among socio-economic factors. This high correlation makes it desirable to distinguish between those variables associated with fatalities and those variables which cause changes in fatalities. This distinction is formalized in the first two steps of our analysis. In the first, 34 socio-economic variables were screened to identify those associated with fatality changes. In the second, the best of these variables were tested for their ability to reduce forecasting error, an empirical test referred to as "statistical causation."

Based on the results of these first two steps, multi-variate models of highway fatalities were developed based on socio-economic factors. These models are based on the hypothesis that socio-economic factors influence fatalities either by influencing the aggregate amount of driving or by influencing the composition of driving, that is, the proportion of driving by certain groups of people, in certain vehicle, under certain conditions. The composition of driving can influence the number of fatalities because the drivers, vehicles, and conditions have different highway fatality risks. If reliable estimates of VMT for the appropriate categories of drivers, vehicles and conditions were available they could be used directly to estimate fatalities. Because they are not available, economic variables which affect the amount and risk distribution of driving were sought. Further, since the economy is widely forecast, knowledge of the influence of economic variables on fatalities is useful itself because it permits forecasts of highway fatalities to be easily developed.

Thus, the socio-economic factors affect the amount of driving or the risk distribution of driving. These two types of influences, aggregate VMT and composition of VMT, are tested separately and then combined into a multivariate model of highway fatalities which can be used to examine the 1982 fatality decline.

In summary, there are five steps in the approach: 1) Screen socio-economic variables for association with fatalities; 2) Test strongly associated variables for "statistical causation;" 3) Develop models which account for the effect of changes in aggregate VMT; 4) Develop models which account for the effect of changes in the risk distribution of driving; and 5) Combine the two types of models and use the results to explore the 1982 fatality decline.

<u>Results</u>

Thirty-four socio-economic variables were screened for association with highway fatalities using a regression procedure that corrects for autocorrelation.*

*Autocorrelation is correlation between successive values in a series. It can lead to inaccurate estimates of regression coefficients and biased significance tests if uncontrolled.

Fourteen of these socio-economic variables were found to be significantly associated with highway fatalities at the 80 percent confidence level. Four of these (the size of the labor force, VMT, gasoline sales, and a derived measure of travel*) were significant at the 95 percent confidence level (see Table E-1). Three of the four, all except the size of the labor force, measure immediate driving activity. This is consistent with the theory that changes in monthly fatalities are most strongly influenced by changes in aggregate monthly driving. The performance of VMT in this screening supports the usefulness of VMT as a measure of total driving activity.

Of the 14 variables that passed the initial screening, eight were found to meet a test of "statistical causation." This test requires that forecasting error is reduced by adding the variable to an explanation of current fatalities which is based only on past values of fatalities. Section 3.0 develops the rationale for this test and shows how it relates to causality.

VMT and four economic variables hypothesized to influence aggregate driving were tested for their ability to forecast highway fatalities over consecutive 12-month periods between 1975 and 1982. Personal income and retail sales were found to forecast fatalities roughly as well as VMT and better than the other two variables, the FRB production index and average weekly earnings of production workers.

"Path analysis" was performed to test empirically the hypothesis that retail sales and personal income influence fatalities through their effect on VMT. This analysis suggested that retail sales and personal income have a substantial influence on the level of VMT, and that retail sales may have an additional effect on fatalities beyond its influence on total VMT. As a result, retail sales and personal income can be interpreted as proxies for VMT in models of highway fatalities.

*Gasoline sales divided by an estimation of MPG.

TABLE E.1

SOCIO-ECONOMIC VARIABLES

FOUND SIGNIFICANT AT THE 80% LEVEL.

	Level of Significance	Statistical Causation
Driving Activity		
VMT (Vehicles Miles Traveled)	95%	YES
GSALES (Gasoline Sales)	95%	YES
GMILES (Gasolilne Sales/MPG)	95%	YES
GAS$ (Gas Price)	80%	NO
CARCOST (Price Index for Private Transportation)	80%	—*
Income and Employment		
UNEMP (Number of Unemployed)	80%	YES
EARN (Average Weekly Earnings of Production Workers)	80%	YES
FRB (Federal Reserve Board production index)	80%	YES
L.IND (Composite of 12 Leading Indicators)	80%	NO
Demographic and Vehicle Fleet		
POP (Population	80%	—*
LF (Labor Force)	95%	YES
NCR (New Car Registrations)	80%	—*
NTR% (New Truck Registrations as a % of New Vehicle Registrations)	90%	—*
Other		
HUDI (Department of Housing and Urban Development Interest Contract rates)	90%	YES

*Not Tested.

Further, retail sales may influence, or be associated with, the risk distribution of total driving (i.e., the proportion of driving done under higher-risk conditions). In an attempt to identify other specific, more direct influences on the risk distribution of total driving, five additional variables were tested for their ability to improve significantly a fatality model based on VMT alone (see Table E-2). Statistical tests could not establish that any of these variables affected the risk distribution of driving.

Analysis of the 1982 decline

The 1982 fatality decline was explained using models based on VMT or its surrogates because these models were found to explain fatalities better than any other models tested. Fatality declines of 6 to 8.5 percent were forecast for 1982 using these models. They started with the December 1981 fatality count and used actual 1982 data on VMT and its proxies. The performance of these models indicates that month-to-month changes in VMT can explain most of the reduction in fatalities and that increases in VMT are not inconsistent with declining fatalities. Furthermore, the month-to-month errors in the fatality estimates were no different in 1982 than they were for the years 1975-82, which indicates that there has been no major change in the relationship between fatalities and VMT in 1982. However, the decline has not been explained fully, and there may be other important influences on the risk distribution of driving that contributed to the 1982 fatality decline.

A detailed examination of the VMT model reveals three factors which help explain the complex relationship between VMT and fatalities. First, fatalities will actually decline unless VMT increases by more than two percent because the fatality rate per vehicle mile is declining. Safer drivers and improvements to roads and vehicles probably account for some of the reduction in the fatality rate. Another cause for the decline is that increases in the number of licensed drivers cause the average mileage per driver to decrease when total VMT increases by less than two percent; and since the safest commuting and shopping trips are not likely to be reduced when the mileage per driver is reduced, safer driving circumstances are substituted for riskier circumstance and the fatality rate declines.

TABLE E.2

INFLUENCES ON THE RISK DISTRIBUTION OF DRIVING (AFTER CONTROLLING FOR AMOUNT OF DRIVING)

Variable	Hypothesized Effect
Teenage Unemployment (YUN)	Higher teenage unemployment causes lower teenage driving and lower fatality rates.
Liquor Sales (LS)	Higher liquor sales causes more drunk driving and higher fatality levels.
National Park Visits (NPM)	Higher National Park visits indicate more vacation travel and higher fatality levels.
Employment (EMP)	Higher employment causes more commuter travel and lower fatality levels.
% New registrations which are trucks (%NTR)	Higher proportion of trucks in the vehicle fleet indicate a higher proportion of heavy truck VMT and higher fatality levels.

Second, VMT growth of more than two percent annually causes total fatalities to increase at a rate of 1.4 percent for each one percent increase in VMT. Again, the driving which is added after the two percent needed to maintain a constant average mileage per driver is likely to be riskier discretionary (non-work) driving which has a higher fatality rate per mile than average.

Third, the monthly distribution of driving has an important influence on total fatalities. A higher share of driving in summer months results in a higher fatality level. This increase reflects more higher-risk discretionary driving in the summer months, with a corresponding higher fatality rate per mile.

Conclusions

- VMT, personal income, and retail sales explain highway fatalities about equally well over the 1975-82 period.
  - The relationship between highway fatalities and these three variables appears to be unchanged in 1982 from the 1976-80 period.
  - The influence of retail sales and personal income on fatalities was found to be partly because of their influence on VMT. The activities of spending and earning involve driving.
  - Path analysis suggests that retail sales also affect the risk distribution of total driving.
- An increase in VMT will not cause an equal percentage increase in fatalities for three reasons:
  - First, a declining fatality rate per vehicle mile causes fatalities to decline if VMT grows by less than two percent. Increases in the number of drivers and consequently the risk distribution of driving and safety improvements probably account for this long-term trend;
  - Second, VMT growth of more than two percent annually increases fatalities at a rate of 1.4 percent for each one percent rise in VMT. This occurs because VMT increases above two percent are concentrated in higher-risk discretionary driving; and
  - Third, growth in VMT that occurs during summer months increases fatalities more than comparable growth during winter months. A

higher share of driving in the summer months reflects more higher-risk discretionary driving (i.e., vacation, entertainment, and recreation-related travel), with correspondingly higher fatality rates per mile.

- Starting with December 1981, a model of month-to-month changes in fatalities that incorporates the three above mentioned VMT effects explains roughly two-thirds of the 1982 fatality decline.
- Other influences on the risk distribution of driving may account for some of the unexplained portion of the 1982 fatality decline. Further study is needed on the impact of variables that influence the distribution of driving.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

In 1982 highway fatalities dropped by more than ten percent while estimated total vehicle miles traveled (VMT) increased by slightly more than one percent. This counter-intuitive trend began in 1981 when fatalities dropped by more than three percent and VMT increased by almost three percent . Since the economy was in a deep recession in 1981 and 1982, with unemployment above ten percent, economic causes for the fatality decrease were suggested.

The objective of this study is to identify socio-economic factors which influence highway fatalities, to test plausible hypotheses on how these factors exert this influence, and to explore the effect of these influences in 1982. Historic data on highway fatalities (monthly fatalities between 1975 and 1982) from NHTSA's Fatal Accident Reporting System (FARS) and time series of socio-economic variables were used to investigate the influence. Because this data is non-experimental data, inferring causality from correlation is not appropriate and more demanding tests of influence are needed. Further, because of strong trends in both the fatality data and the socio-economic data series and because of strong auto-correlation in these data, special statistical procedures are needed in order to interpret the levels of significance of the estimated coefficients. Section 3.0 discusses the methods used in this study to deal with these problems.

Section 2.0 presents a brief review of some past studies of the socio-economic influences on fatalities and concludes with a broad discussion of categories of socio-economic variables and how they are hypothesized to influence fatalities.

Section 4.0 presents the results of screening 34 socio-economic variables using the methods described in Section 3.0. This section ends with a comparison of the forecasting performance of the strongest of the 34 variables.

In Section 5.0, multivariate models are hypothesized and the forecasting performance of these models evaluated. This section also contains a test of the hypothesized path or chain of influence of the variables in the multivariate models.

Finally in Section 6.0, the best multivariate models are used to estimate fatalities in 1982 and the behavior captured by the models is analyzed.

2.0 Socio-Economic Influences on Fatalities

In this section, hypotheses about the influences of socio-economic variables on fatalities are discussed first from the perspective of previous work and then in conjunction with this study. It will be helpful to categorize the socio-economic variables as follows:

Driving Activity
Driving Cost
Income and Employment
Demographic (Population and Motor Vehicle Fleet)
Other

2.1 Previous Work

A number of previous studies have used time-series techniques to examine the relationships between highway fatalities and socio-economic variables. These works were useful in suggesting hypotheses and variables for testing with the methods used in this study. Four specific works [1, 2, 3, 4] were reviewed but the models contained were not appropriate for this study because:

1. The data used in estimation ranged up to the year 1975 and thus omitted the 1980-82 fatality decline.

2. The models were not based on monthly data. Presently the FARS database allows for analysis of eight years of monthly observations on highway fatalities.

3. Many of the models normalized highway fatalities (the dependent variable), i.e., fatalities per capita, fatalities per VMT, fatalities per constant risk. This type of normalization is not used in this study because the variables used in normalizing may have important explanatory power and because if the normalizing variables are correlated with fatalities or with independent variables, statistical testing of the model becomes difficult.

4. Many of the models do not control for autocorrelation, and autocorrelation in the highway fatality and independent variable series will cause biases in statistical testing.

5. Groups of variables included in some models move together (multicollinearity) and it is difficult to interpret the individual effect of a variable.

6. A number of potentially important socio-economic variables are not tested.

Table 2.1 presents the variables used by category for the four previous works. Since the methods of these works often involve searching for an optimum specification, a variable may appear in several different equations.

Driving Activity

Land and McMilien, in explaining fatalities, use speed as the best control of exposure to risk. Speed is measured as the percent of vehicles recorded as exceeding 60 miles per hour on straight sections of main rural highways. It is assumed that as speed increases so will risk. When Land and McMilien estimate amodel to explain highway fàtalities with the variables of age structure of drivers, highway crowding and miles driven per vehicle, these variables are significant only

TABLE 2.1 SOCIO-ECONOMIC VARIABLES USED IN FOUR PREVIOUS WORKS

	L + M[1]	PELTZMAN[2]	ABT[3]	JOKSCH[4]
DRIVING ACTIVITY				
Speed	X	X	X	
VMT	X		X	
DRIVING COST				
Gas Price			X	
Accident Costs		X		
INCOME AND EMPLOYMENT				
Unemployment Rate	X			
Per Capita Earned Income		X		
FRB			X	X
DEMOGRAPHIC (Population and Motor Vehicle Fleet)				
15 to 24 year-olds as percentage of total population	X	X		
Car Registrations	X			X
OTHER				
Linear Trend		X		
Liquor Consumption		X		

when speed is not included. Thus, they conclude that speed is the crucial risk exposure variable which serves as a proxy for the other three variables.

Peltzman uses speed to test the hypothesis that increased speed causes increased fatalities. In both his time-series and cross-sectional models the speed term is highly significant. Peltzman places the interpretation on speed that it is one way of satisfying the demand for risky driving. This demand increases as real income increases because the time spent in travel is more valuable. Economizing by driving faster is especially likely, according to Peltzman, when time spent driving reduces income earning possibilities.

In the ARIMA models of the Abt work there is no causal interpretation given to the speed variable — only that the speed and fatality series move together. No causal interpretation is made for any variable included in the Abt work. Their purpose was to find variables that tracked the trend of the fatality series.

Land and McMilien consider VMT per registered vehicle as an exposure index. Ideally, they would like to normalize not with registered vehicles but with registered drivers if that data were available for the 1946-72 period. As mentioned above, speed is used as a partial proxy to measure the risk associated with increased driving. As well, Land and McMilien find changes in driving are associated with the unemployment rate. During an economic downturn there should be less commercial vehicle travel and possibly a decline in non-commercial travel also.

Driving Cost

Peltzman argues that driving produces accidents and that these accidents have costs to the involved drivers. As these costs increase there should be a disincentive to having accidents. These costs are measured with an index of medical and repair costs multiplied by an insurance loading factor, the ratio of the current year's premiums to last year's benefits paid. As the loading factor increases, so will the cost borne by the driver.

Income and Employment

Peltzman uses various income measures to make the interpretation that short-run (transitory) increases in income are associated with higher fatality and accident rates in the short term because increases in especially wage income causes demand for more "driving intensity" as through speed (discussed above). In the long term, Peltzman believes permanent increases in income cause a demand for more vehicle safety and thus tend to reduce the death rate. This long term effect is captured in a linear trend term.

Joksch uses the FRB Index of Industrial Production in making year-to-year predictions (as distinguished from predicting just the fatality trend) of fatalities, but no causal reason for this relationship is given other than there may be some common factor which affects both fatalities and the FRB index.

Demographic (Population and Motor Vehicle Fleet)

Land and McMilien use the number of males 15-to-24 years old as a percentage of the total population in a beginning specification because this age-sex cohort is proportionately the most involved in fatalities. However, as mentioned above, this variable is replaced by speed in a final parsimonious specification.

Peltzman also uses the age composition variable and concludes from its significance that youthful drivers have a different outlook or "taste" for risk, although he points out that there is some evidence of increased blood alcohol concentrations in this age group.

Land and McMilien use the number of registered vehicles divided by the number of urban and rural miles of highway as an index of crowding which is also replaced by speed in a subsequent model.

Joksch, in constructing a trend model of passenger fatalities, groups the number of passenger cars into three age categories: (1) current model year; (2) 1-to-3 years old; and (3) four years or older. This grouping of cars by age performed better in predicting the fatality trend than other measures tried: VMT, truck registrations, FRB production index, speed, and a time factor. As Joksch points out, the model's results are dependent on a stable relationship between the age distribution of the passenger car stock and if a unique event should cause a large drop in new car sales the model may not forecast well.

Other

Peltzman uses a trend term to capture secular forces which in the long run reduce the fatality rate. As mentioned above, Peltzman believes that increases in real income cause demand for safer vehicle design. Other factors represented by the trend are: improvements in highway design, driver skill, health care, and vehicle maintenance.

To test what effects alcohol has on fatalities, Peltzman includes in his model the consumption of distilled spirits per person 15 years of age and older. The variable is significant and has the correct sign (directional influence).

2.2 The Influences on Fatalities

In this section, the socio-economic variables hypothesized to influence highway fatalities are further described in terms of **a priori** expectations. Again, because there are a large number of these variables, categories are useful in discussing the concept common to a group of variables. These categories were presented at the beginning of Section 2.0.

Driving Activity

The fundamental influence on fatalities is the amount of driving done by specific driver groups under specific conditions. Fatalities result from driving and the total number of fatalities from this driving is the result of the inherent risk. Unfortunately VMT by type of driver is not available for the monthly time-series framework of this study, nor was there data available on type of driving other than urban-rural designations. Thus only aggregate measures of driving can be used, the most popular being VMT* (vehicle miles traveled). Other, less direct measures can be used such as monthly gasoline sales and total sales of the products at gasoline service stations. These latter variables might be used to test whether published VMT contains measurement error or sampling variation.

Driving Cost

The real price of gasoline and an index of automobile costs affects the level of driving activity. Increases in these measures (all other things remaining the same) should reduce driving and hence accidents. Further, these variables may affect the type of driving; less discretionary driving** results when prices increase. Discretionary driving, which is more likely to be at night or on weekends and may involve the use of alcohol, is more risky.

Price increases may affect the distribution of drivers as well, since different drivers have different sensitivities to price changes. A full specification of the cost concept might include a relative cost variable such as the gasoline price divided by the Consumer Price Index (CPI). This would be a measure of the desirability of substitution between driving and other goods and services.

Income and Employment

Income and employment variables also serve as less direct measures of driving activity. High income, all other things being equal, should increase the amount of driving and hence, increase fatalities. Because at any one time there is a fixed level of necessary driving, sharp increases or decreases in income and employment

*Yearly VMT is published in Highway Statistics by type of vehicle and monthly VMT is contained in "Traffic Volume Trends."

**Discretionary travel is defined as nonwork travel. There is empirical evidence[5] that discretionary travel is more sensitive to gasoline price changes than work travel. That is, discretionary travel is more price elastic.

Demographic (Population and Motor Vehicle Fleet)

variables should cause changes in the riskier discretionary driving. There are many temporal measures of income and employment and indirect measures such as the Federal Reserve Board (FRB) production index. The Department of Commerce's Composite Index of 12 Leading Indicators may provide a more distant outlook on economic conditions and hence fatalities.

The population and the number of people in the labor force at various times are thought to have positive relationships with travel (and indirect measures of that travel) and hence fatalities. The number of people who are registered drivers or the number of vehicles owned by the population should also exert positive influences on travel and fatalities.

Other

Several variables which did not fall naturally into the above categories are worth examining. The CPI by itself would measure increases or decreases in goods and services which may compete with spending on travel. It may also serve to measure expectations of future economic conditions. The level of the interest rate may serve as a proxy for driving costs, expectations of future conditions, and the wealth of consumers in that an increase (decrease) in the interest rate decreases (increases) the value of existing financial assets (bonds, stocks). Finally, liquor sales may serve to measure the important issue of alcohol involvement and the current trend of that involvement with fatalities.

In summary, it is hypothesized that socio-economic variables: (1) affect the amount of driving; (2) affect the distribution of driving between discretionary and non-discretionary, each with a different risk; and (3) affect the proportion of total driving by groups of drivers with differing total accident risk. As economic conditions change, these three aspects of driving change and produce different levels of highway fatalities.

3.0 Statistical Procedures

The search for cause and effect among variables is one of the major goals in scientific research. However, when it is not possible to conduct a controlled experiment, it often becomes difficult to produce convincing evidence that a cause and effect relationship actually exists. Inevitably this is the case in most social, economic, and transportation research.

The common approach adopted by many social and economic researchers in analyzing non-experimental (observed) data is summarized below:

(1) On the basis of incomplete prior knowledge, a single-equation regression model is specified. By so specifying, an assumption has been made that causality flows from the independent variables to the dependent variable: an asymmetrical causal relationship and the direction of that causality is taken to be known.

(2) The model so specified with a dependent variable, Y, and a set of independent variables, X_i, is then estimated with a standard regression package to test the equation and coefficients of the independent variables for statistical significance. Then, interpretations of cause and effect are made or implied by the researcher.

The basic weakness of this approach is that too much faith is placed on prior knowledge of the causal relationship in specifying the model. If prior knowledge is not complete or correct, the significance of the regression coefficients (as partial correlations with the dependent variable) and the significance of the independent variables (as a whole in explaining variation in the dependent variable) cannot provide conclusive evidence of a causal relationship. This is because such significance is also compatible with the hypotheses that the variables included in the equation may have the reverse direction of causality than the one specified, or are related to omitted variables.* If they are related to the omitted variables, the included variables could simply move together, the real cause of this movement being an omitted variable. As an example, in a system containing two variables, X and Y, a "weak causal ordering" relationship is defined as: (1) X may or may not cause Y; and (2) Y does not cause X. The classical example of weak causal ordering is that a person's sex (X) may or may not influence political preference (Y) but certainly political preference cannot influence sex. Testing the regression coefficient for significance may provide some evidence of a causal relationship and the direction of causation. However, if prior knowledge is not complete or correct about the relationship between X and Y, then the common approach will not provide convincing evidence. When working with non-experimental data, it is often the case that precise knowledge of the weak causal ordering among the variables of interest is unknown.

Another problem has been observed in the common approach for determining causality. Time-series data usually are autocorrelated. That is, values in one time period are similar to values in a previous period. For example, high values in one period usually are associated with high values in neighboring periods. There can be many different patterns of autocorrelation, but the basic problem is lack of independence of successive values. This lack of independence can cause difficulties in identifying the historical relationships between X and Y. If the two series are causally related, then the covariance between the series will not be zero. The series will move together because of the causal relationship. But if the series are not causally related and one or both of the series are autocorrelated, then the covariance will still not be zero. Hence, if the series are autocorrelated, it is not possible to distinguish between a causal relationship and autocorrelation in the series. A causal relationship may appear to exist in the sample period where none really is present.

Autocorrelation causes another more technical problem as well. Autocorrelation in a variable can cause autocorrelation of the regression residuals* if the Ordinary

*Examples of these two types of errors are: (1) Correlation between employee education and income does not necessarily mean that higher income causes better education; (2) Correlation between accidental drownings and ice cream sales might be strong, but an omitted variable, weather, causes both independent activities.

Least Square (OLS) procedure is used to estimate the regression equation. This bias will cause errors in accepting or rejecting hypotheses about the regression coefficients.

To avoid these shortcomings of the common approach, it is necessary to develop a sound statistical approach for examining causal relationships among a set of variables. This need has led to a precise statistical definition of "causality." The statistical definition refers to the idea of predictability, X statistically "causes" Y if the X series is able reliably to predict future values of the Y series. This statistical concept of "causality" is useful because it can be empirically tested but it is much looser than causality in the ordinary usage of the term.

The concept of causality as applied to time-series data was first formalized by Granger (1969)(6). Others have developed methods of implementing and testing for causality in time-series data. The next two sections will discuss the concept and give alternative tests of causality with non-experimental data.

3.1 A Statistical Concept of Causality

Granger (1969) has proposed a statistical concept of causality for time-series data. The major merit of his concept is that it can be tested empirically. His idea of causality is relevant only for stochastic time-series variables (variables with random components) and is based upon prediction. A stochastic variable X is said to statistically cause another stochastic variable Y if knowledge of X improves our ability to predict Y, given that all other information about Y has been used.

*The residual is the unexplained portion of the variation in one series (dependent variable) which is not explained by one or more other series (independent variables) within the framework of a statistical model. For proper statistical testing of the model there should be no discernible pattern in these residuals other than randomness. Autocorrelation in the residuals means that successive values are correlated. Autocorrelation of the regression residuals will bias the standard error of the regression equation.

More formally, let:

U_t = all knowledge of the universe up to and including time t-1;

X'_t = all past values of X up to and including time t-1;

X_t = all values of X up to and including time t;

$\sigma^2(Y/U)$ = minimum predictive error variance of Y_t given U_t;

$\sigma^2(Y/U - X')$ = minimum predictive error variance of Y_t given U_t apart from X'_t;

$\sigma^2(Y/U, X)$ = minimum predictive error variance of Y_t given U_t and X_t.

Granger defines:

(1) past X causes current Y if

$\sigma^2(Y/U) < \sigma^2(Y/U - X')$; and

(2) X causes Y contemporaneously if

$\sigma^2(Y/U, X) < \sigma^2(Y/U)$.

Causality from Y to X is defined in the same manner. If X both causes Y and is caused by Y, there is feedback or bi-directional causality between the two variables.

Granger's definitions appear to be basically consistent with the common sense notions of causation.[7] For example, he incorporates the common sense idea that future events cannot cause current or past events. The definition (1) is also consistent with the idea that, in order for past X to cause Y_t, the past values of X must have some effect upon Y_t which is independent of all other forces which affect Y_t (either directly or through past X).

However, Granger's definitions are not empirically applicable except under ideal conditions which are probably never met. It seems unlikely, for example, that we are ever going to be able to predict Y_t on the basis of all knowledge of the universe in the past, U_t, and then to contrast this with predictions based upon U_t apart from X'_t. We do not have all knowledge about U_t.

Granger is well aware of this difficulty and has presented what might be described as "relative" or "constrained" definitions of causation which are empirically applicable. Given an information set $A = A_{t-j}$; $j = 0, 1, 2, \ldots$ which includes at least the time-series variables Y and X, Granger defines:

(3) past X' causes current Y relative to the information set A if $\sigma^2(Y/A) < \sigma^2(Y/A - X')$

(4) X causes Y contemporaneously relative to the information set A if $\sigma^2(Y/A, X) < \sigma^2(Y/A)$.

We find definitions (3) and (4) to be of limited usefulness because we have no common sense understanding of what it means to say that "one variable causes another relative to some information set." We prefer to stick with Granger's original definitions and in empirical work to employ what are in the nature of necessary (but not sufficient) conditions for one variable to cause another, for example,

(5) if past X causes current Y, then

$$\sigma^2(Y/A) < \sigma^2(Y/A - X')$$

(6) if X causes Y contemporaneously, then

$$\sigma^2(Y/A,X) < \sigma^2(Y/A).$$

The use of conditions (5) and (6), makes it very clear that our empirical tests are not capable of establishing that one variable actually causes another. They can only tell us when one variable does not cause another. Suppose we find that σ^2 $(Y/A) < \sigma^2 (Y/A-X')$. We know that this is consistent with the hypothesis that X causes Y and also with the hypothesis that Z (an omitted variable) causes X and Z causes Y but X does not cause Y. However, if we find that $\sigma^2 (Y/A) \geq \sigma^2 (Y/A-X')$ we see that X does not reduce the prediction error and does not cause Y.

3.2 Tests of Causality

This section presents two empirical tests of causality consistent with Granger's ideas:

(1) a modified Yeats procedure[8]; and
(2) an autoregressive modeling approach.

The selection of these procedures is guided by two considerations:

(1) the computations and statistical testing can easily be done by existing statistical software; and

(2) the statistical tests and methodologies are understandable to the general researcher.

3.2.1 A Modified Yeats Procedure

Yeats (1972)[9] proposes a simple regression procedure to identify one time series as a leading indicator of another. The procedure is as follows:

$$(7) \Delta Y_t = a + b_n \Delta X_{t-n} + e_t$$

where Δ denotes the change or first difference, a and b_n are the Ordinary Least Squares estimates, and e_t is the random error term.

If X_t is in fact a leading indicator (X causes Y), it should, with some degree of regularity, change in a consistent direction before the Y_t series. If the estimated b_n coefficient ($n = 1, 2, \ldots, L$) is statistically significant with the appropriate directional sign, there is some evidence that X_t leads the Y_t series by n periods. That is, the X_t causes Y_t in the sense of prediction. Similarly, we can regress X_t on Y_{t-j} ($j = 1, 2, \ldots, L$) to test whether Y_t is a leading indicator of X_t.

In this study, the Yeats procedure is employed to evaluate leading properties of selected socio-economic variables with respect to the total fatalities series. The procedure , however, is slightly modified and the model is:

$$(8) \Delta \log Y_t = a + b_n \Delta(\log X_{t-n}) + \sum^{11} d_i \, SD_i + U_t$$

where: $\Delta \log Y_t$ is the logarithmic first-difference of the total fatality series adjusted for working day/trading day variation* ($\log Y_t - \log Y_{t-1}$ or equivalently $\log (Y_t/Y_{t-1})$)

a is the intercept term which picks up the effect of any time trend present in the model.

b_n is the coefficient of the socio-economic variable being tested at a lag of n months for n=0, 1, ... 12.

$\Delta(\log X_{t-n})$ is the logarithmic first-difference of the socio-economic series being tested, either the current period (n=0), $\log X_t - \log X_{t-1}$, or a lagged period, $\log X_{t-n} - \log X_{t-n-1}$.

d_i is the set of coefficients for the 11 seasonal dummy terms used to adjust Y and X for seasonal variation.

* For a complete description of this procedure see reference (10) as well as the last two paragraphs of Section 3.2.1.

SD_{it} are a set of 11 seasonal dummies (i=1, ..., 11). Each dummy vector is coded:

SD_{it} = 1 if the fatality count is observed in month i of the year.
-1 if the fatality count is observed in December.
0 otherwise.

The coding of December as "-1" for all eleven SD_i imposes the restriction that the values of all the twelve seasonal effects (adjustments) sum to zero. Thus, the December value can be derived by algebraically summing (the coefficients) of the first eleven months and reversing the sign. With this type of specification, the seasonal effects measure deviations from an average month.

U_t is the stochastic error term of the equation. U_t should be stationary and hence have a constant variance.

The formal Yeats model has been modified in several ways. First, the first differences of the natural logarithm is used because it is expected that percentage, and not constant, change in the variables should be related. Second, seasonal adjustments have been made in the regression equation because monthly data are being used. By not pre-adjusting each series for seasonality before estimating the equation, proper degrees-of-freedom adjustments are being made in calculating the regression statistics. Further, if one seasonal adjustment procedure is used on one variable and another type of procedure is used on a second variable, including both these variables in the same regression equation can produce biased estimates of the coefficients. It is often the case with published data that different seasonal adjustment methods have been used and it is often difficult to determine what these methods are. Including the seasonal adjustment in the regression equation protects the coefficients from this bias.

It should be mentioned that using the first difference is an important aspect of the procedure as it will often eliminate any trend that may appear in either the fatality or socio-economic series and thus reduce the possibility that regression results will only be finding proportionality among two series with a common trend. Using first differences also provides a beginning correction procedure for autocorrelation as the differences are less likely to be correlated than successive absolute amounts. If first order autocorrelation is observed in the residuals, as revealed by the Durbin-Watson statistics, then a Generalized Least Squares (GLS) procedure is used which corrects for this autocorrelation. Ideally, the residuals should be formally tested for different structures of autocorrelation, most notably autocorrelation with a 12-month period.

The monthly highway fatality series was taken from the FARS database for 1975-82 as of February 1983.* Monthly data was used to give the maximum number of observations (96) available from this source and to allow determination of the seasonal pattern over time. These counts were adjusted for working day/trading day variation because it was observed that the fatality rate per day is higher on weekends and holidays than for weekdays, and even that there is variation for different days of the week. Thus, working day/trading day adjustments standardize the monthly fatalities for a period of similar day exposure.** This adjustment was done prior to the Yeats procedure. The adjustment should lessen autocorrelation between successive monthly fatality counts because after adjustment high fatality months with a proportionality high number of weekends will not be followed by low fatality months with a low number of weekends, or vice-versa.

The working day/trading day adjustment is done by regressing the monthly fatality counts from the FARS database on variables representing the number of weekdays and weekends while controlling for trend and seasonality. The coefficients of the weekday and weekend variables are then used to adjust the fatality series.

*At that time the 1982 file was incomplete and it was necessary to estimate monthly fatality totals for 1982 from this incomplete file. This estimate proved to be high by 549 fatalities in total for 1982.

** For a complete description of this procedure see Reference (9).

3.2.2 Autoregressive Modeling Approach

The second test examines the statistical causal ordering between two variables and was suggested by Granger (1969). To test for statistical causality, it is assumed that the information relevant to the prediction of the respective variables is contained solely in the data series X and Y. This direct regression procedure is described as follows:

(1) If a system contains two stationary time series X and Y, a pair of regression equations can be specified as follows:

$$(9)\quad Y_t = \sum_{t=1}^{n} a_i X_{t-i} + \sum_{t=1}^{m} b_j Y_{t-j} + e_t$$

$$(10)\quad X_t = \sum_{t=1}^{H} c_i X_{t-i} + \sum_{t=1}^{L} d_j Y_{t-j} + f_t$$

It is assumed that e_t and f_t are uncorrelated error terms. OLS is used to estimate the parameters and related statistics of equations (9) and (10).

(2) Unidirectional causation from X to Y is implied if the estimated coefficients on the lagged X variable in equation (9) are statistically different from zero as a group and the estimated coefficients on the lagged Y variable in equation (10) are not statistically different from zero. Conversely, unidirectional causation from Y to X exists if the lagged X coefficients in equation (9) are not statistically different from zero and the lagged Y's in equation (10) are statistically non-zero as a group. Feedback is suggested when the X and Y coefficients are statistically different from zero in both equations (9) and (10). Independence occurs when X_t and Y_t coefficients are not significant in equation (9) and (10).

If the X_t and Y_t series are homogeneous nonstationary* series, then taking the first difference of X_t and Y_t will usually reduce each to a stationary series. Hence, in this case Y_t and X_t in equations (1) and (2) are replaced by $\Delta Y_t = Y_t - Y_{t-1}$ and $\Delta X_t = X_t - X_{t-1}$.

The major weakness of the general approach is that it requires researchers to select n and m arbitrarily in equation (9) (and K and L in equation (10)). It is well known that if the selected m is smaller than the true m, then the OLS estimates of the parameters of equation (9) are biased, because the assumptions of independence of e_t and f_t are violated due to the misspecification of the models. On the other hand, in the case where the number of included lagged Y_t terms is larger than the required m, the estimated variance of the residuals will be unduly increased. Hence, it will weaken the power of the F-test. Clearly, an operational test for selecting m and n is required in implementing these procedures when researchers do not have prior knowledge of the appropriate order of m and n.

Hsiao (1979)[7] proposed a stepwise procedure based on Akaike's final prediction error (FPE) and the statistical concept of causality to fit multiple autoregressions.

The FPE is defined as a mean squared prediction error.

$$\text{FPE of } Y_t = E\,(Y_t - \hat{Y}_t)^2$$

where $\hat{Y}_t$ is the predictor of Y_t

$$\hat{Y}_t = \sum^{n} a_i X_{t-i} + \sum^{m} b_j Y_{t-j}$$

* A nonstationary series has a mean and variance which do not change with time. If the series can be reduced to one having a mean and variance which do not change with time by taking the first difference one or more times, it is said to be a homogeneous nonstationary series. The number of first differences needed is the order of homogeneity of the series.

The m and n denote the order of lags in b_j and a_i. a_i, and b_j are least squares estimates. The FPE in this case is given by

$$FPE\ (m, n) = \left[\frac{T + m + n - 1}{T - m - n - 1}\right] * \left[\sum^{T} (Y_t - \hat{Y}_t)^2 / T\right]$$

The first factor can be considered as a measure of estimator error and the second factor as a measure of the modeling error. The criteria tries to balance the risk resulting from bias when a lower order lag is selected and the risk resulting from the increase of variance when a higher order is selected by choosing the specification that gives the smallest FPE.

Combining the definition of statistical causality and FPE criterion, Hsiao suggests the following sequential procedure for identifying the best lags in equations (9) and (10):

(1) Take Y as the dependent variable. Determine the best lag (m_1) for the one-dimensional autogressive equation for Y by minimizing the FPE criterion.

(2) Assume X as a causal variable that affects the outcome of Y. Use the FPE criterion to determine the best lag of X, say n_1, assuming the lag of Y is the one specified in step 1, say m_1.

(3) Check whether the best lag of Y (m_1) might be affected by the X lag n_1 by fixing the X lag at n_1 and searching for the new m_2 with the lowest FPE, if m_2 equals the old m_1 then stop. Otherwise repeat the process holding m_2 fixed and searching for a new n_2.

(4) Compare the smallest FPE's of step 1 and step 3 and if the former FPE(m,0) is less than the latter FPE (m,n), an unvariate autoregressive model for Y is used. If the converse is true, one can say X causes Y and the optimal model for predicting Y is the one, including m lagged Y and n lagged X.

(5) Repeat steps 1 to 4 for the X series and treat Y as the causal variable.

3.2.3 Tests of Forecasting Performance

After the optimal lag structure for Y and X — the m and n — is selected, a test can be performed to select which X variable performs the best in actual forecasting. This test estimates actual historical values using parameters obtained from past data, not including the historical values that will be forecast, to compare the accuracy of the alternative models. This method is a split data technique and is another criterion for use in variable selection. A high $\bar{R}^2$ in model estimation does not necessarily imply that the model will predict well outside the sample period. The data are split into two parts: the first part is used to estimate the model with a specific independent variable; the second part of the data, which might be the last 24 months, is reserved to test the forecasting performance of the model. Hence, the values of the independent variables used to forecast accidents are known. Therefore, the forecasting errors can be attributed to the specification of models. A forecast, so obtained, is known as an ex-post forecast.

This forecasting test was performed using variables which showed statistical significance in the Yeats procedure. The optimal lag structure was determined using the autoregressive modeling process. The results and interpretations are presented in Section 4.0.

3.3 Path Analysis

Where previous sections explain methods for determining causal relationships between highway fatalities (a dependent variable) and a socio-economic variable (on independent variable), this section explains path analysis, a procedure for determining how several independent variables may interact in influencing a dependent variable. For instance, if both VMT and aggregate personal income are found to be causally related to fatalities, path anlaysis can test the hypothesis that personal income acts to establish the level of VMT which in turn determines the number of fatalities: that is, a "path" of causation exists from personal income to fatalities via VMT.

Path analysis was suggested by Wright (1934)[11] and used by Tukey (1954)[12] and many others in social science research (see Ducan (1966)).[13] It is not a statistical procedure, per se, but a way of using existing statistical methods in the framework of a quantitative model to analyze data. In the case of this study, multiple regression techniques are used to test a chain of causality that is hypothesized to exist among socio-economic variables and fatalities. If a socio-economic variable influences fatalities it may do so directly and/or indirectly by influencing another socio-economic variable which in turn influences fatalities. If both paths exist, path analysis decomposes the influence quantitatively into the direct and indirect effects so that a relative comparison of each can be made.

The first step in path anlaysis is to define a set of relevant variables and a definite causal ordering among these variables from theoretical considerations and/or past empirical work. The causal ordering is such that for a group of variables (X_1, X_2, ..., X_k) those with a higher subscript cannot influence those with a lower. That is X_1 may affect X_3, as may X_2, but X_3 cannot affect X_1 or X_2. Choosing three variables for ease in explanation, this relationship can be expressed in a recursive equation system:

$$X_2 = P_{21}\, X_1 + E_2$$
$$X_3 = P_{31}\, X_1 + P_{32}\, X_2 + E_3$$

The X_i's are variables measured in standardized units* and thus the path coefficients (P_i's) are also standardized** and comparable within the same equation. The E_i's are the residual portion of the equation which cannot be explained by the variation in the X_i's. Sample data is used to estimate the values

* $X_i = (\breve{X}_i - \bar{X}) / S_i$

**See Section 5.2 for the discussion of Beta Coefficients.

of the path coefficients. Ordinary Least Squares estimation may be used as long as the residuals, E_2 and E_3, are not autocorrelated. If that is the case, Generalized Least Squares should be used as a correction procedure. If either of these single equation estimating techniques are to be used, E_2 and E_3 must not be correlated across the equations. If E_2 and E_3 are correlated then a system estimating technique should be used.

Given that the path coefficients (P_i's) have been estimated by the appropriate technique, causal connections among several variables can be evaluated and direct and indirect effects apportioned.

In the above system of equations, the direct effect of X_1 upon X_3 is estimated by $\hat{P}_{31}$. The indirect effect of X_1 upon X_3 via X_2 is estimated by $\hat{P}_{21} * \hat{P}_{31}$. Clearly, if any of the relevant P_i coefficients are not statistically different from zero by a t-test, the hypothesized idea of causal ordering is not supported by the sample data. It is also important that the total variation explained by the equation be significant and high so that the unexplained variation in the dependent variable do not dominate. The amount of explained variation is measured by the corrected R^2:

$$R^2 = (1 - \Sigma\hat{e}_i^2/N\text{-}k) * (N\text{-}1/\Sigma(Y_i - \bar{Y})^2)$$

where: $\hat{e}$ = individual residuals of E_2 or E_3

N = number of observations

k = number of independent variables

Y = dependent variable, X_2 or X_3

The significance of the equation is measured by the F-statistic:

$$F_{k-1,\ N-k} = (R^2/1-R^2) * (N\text{-}k/k\text{-}1)$$

where: $R^2 = 1 - \Sigma\hat{e}_i^2/\Sigma(Y_i - Y)^2$

4.0 Results

This section presents the results of applying the statistical procedures described in Section 3.0 to a large set of potentially important socio-economic variables. The first subsection, 4.1, presents the results of applying the modified Yeats procedure to screen the large set of variables down to a manageable set of more promising variables.

Section 4.2 presents the results of the autoregressive modeling on this smaller set of variables. This modeling identifies optimal lags for the dependent and independent variables and provides the most direct test of statistical causality.

Finally Section 4.3 presents the results of testing the forecasting performance of the models identified in Section 4.2.

4.1 Results of Applying the Modified Yeats Procedure

4.1.1 Method

The modified Yeats procedure was used to screen socio-economic variables which were believed to have an influence on highway fatalities. The modified Yeats procedure was used for the screening because it could be easily applied and it controls for seasonality and trend. These variables were inserted one at a time on the independent-variable side of the modified Yeats equation. The dependent variable of the equation is always the logarithmic first difference of the present month fatality level and the previous month level. This functional form is equivalent to the percentage change of fatalities for a specific month over the previous month.

The first specification of the socio-economic variable to be tested is also the percentage change in the current month from the previous month. After this change is tested in the Yeats regression equation and the value of the regression coefficient and t-statistic noted for concurrent association, the time frame of the variable is changed to the percentage change in the previous month over the next previous month—a time lag of one period. The concurrent variable is dropped from the regression equation so that at any one time there is only one variable in the equation representing the effect of the socio-economic variable and not a lagged structure of several terms. Each independent variable was tested in the modified Yeats procedure with lags of one to twelve months.

Table 4.1 gives the abbreviations and definitions used in the remainder of the report. Appendix A presents the source and the frequency of update for each variable.

4.1.2 Summary of Results

For each variable tested, Table 4.2 gives the t-statistic and regression coefficient for the strongest lag (usually one month) which had the expected sign. Note that only lags of one month or more were considered in Table 4.2 because relationships where past X values influence future Y values can only be interpreted as X influences Y. The time ordering precludes future Y from influencing past X. The coefficient and t-statistics values for the current period and for all lags up to period twelve are in Appendix B. The strongest lag is usually one month. This was judged by the t-statistic for the coefficient, the sign of the coefficient and the stability of the sign in adjacent months. As the data in Appendix B reveal, a one month lag did not always have the largest t-statistic. Larger t values were judged to be weaker particularly when these higher values were observed for lags of more that 6 months.* T-statistic values equal to, or greater than, 1.30, 1.67, and 2.00 represent probabilities of being due to chance of .2, .1, and .05 (significant at the 80, 90, and 95 percent levels) respectively for the degrees of freedom in this sample.

It should be noted when examining t-statistics that a value of 4.00 is not twice as significant as one of 2.00—the t-statistic does not have interval properties.

*Reason and prior knowledge are used to rule out lags with significant t-statistics at more than six months. Large t-statistics after six months may be caused by sampling variation or an unusual observation.

Both values are significant and strict statistical interpretations beyond this are difficult. However, it has been observed in empirical work that high t-statistics tend to lead to more stable relationships and often give better forecasts—but there are many exceptions to this and caution is warranted. Further, when a large number of t-tests are computed an outcome which has slightly less than a probability of .2 of being due to chance is not particularly compelling.

In Table 4.2, all variables which are significant at the 80 percent level or more have the expected direction of influence (sign).

Miles-per-gallon (fuel efficiency), car registrations, and licensed drivers were only available on an annual basis and monthly values were interpolated. These variables were not expected to be significant because the cycles are smoothed by this interpolation.

The 34 variables tested in the Yeats procedure have been functionally classified into five categories to facilitate discussion: (1) Driving Activity, (2) Driving Cost, (3) Income and Employment, (4) Demographic (Population and Motor Vehicle Fleet), and (5) Other.

Variables in the Driving Activity classification do well both relatively and absolutely with the exception of service station sales (STATSLS). Gasoline sales (GSALES) and gasoline sales divided by the interpolated miles-per-gallon variable (GMILES) do about as well as VMT. All three are significant at the 95 percent level.

Driving Cost variables—the real gasoline price (GAS$) and an index of car costs (CARCOST)—are significant at the 80 percent level and have the same t-statistic values.

The total number of unemployed (UNEMP), real average weekly earnings per production worker (EARN), and the Federal Reserve Board Index of Production (FRB) are also all significant at the 80 percent level lagged one month. An index of leading economic indicators (L.IND) is significant at the 80 percent level with a four month lag.

Table 4.1 Variable Definitions

DRIVING ACTIVITY	
VMT	Vehicle miles traveled (billions).
GSALES	Wholesale gross gallons of gasoline reported to state tax agencies (thousands of gallons).
GMILES	GSALES divided by miles per gallon. Annual miles per gallon values are interpolated to monthly values.
STATSLS	Estimated Monthly Retail Sales for Gasoline Service Stations (millions of 72$).
MPG	VMT divided by average monthly gasoline price (GAS).
DRIVING COST	
GAS$	Average gasoline price gallon (72$).
CARCOST	"Consumer Price Index for Private Transportation: U.S. City Average" 1967 = 100.
INCOME AND EMPLOYMENT	
UNEMP	Number of persons unemployed (thousands).
UNEMP%	UNEMP divided by labor force (LF).
YUNEMP	Number of unemployed, age 16-19, both sexes (thousands).
DUNEMP	Average (mean) duration of unemployment (weeks).

EMP Number of employed (thousands).

EMP% Employed as a percentage of working age population.

DI Disposable income. Seasonally adjusted at annual rates (billions of 72$).

DIPC Disposable income per capita. Seasonally adjusted at annual rates (billions of 72$).

PI Personal income. Seasonally adjusted at annual rates (billions of 72$).

EARN Gross average weekly earnings for production or nonsupervisory workers on non-argricultural payrolls (72$).

RS Total Retail Trade (millions of 72$).

FRB Federal Reserve Board Index of Quantity Output (Industrial Production) 1967 = 100.

C.IND Composite Index of 4 Roughly Coincident Indicators.

L.IND Composite Index of 12 Leading Indicators.

<u>DEMOGRAPHIC</u> (Population and Motor Vehicle Fleet)

POP Total noninstitutional population (thousands).

LF Labor force (thousands).

DRIVERS Registered drivers. Monthly values are interpolated from annual values (millions).

CARREG	Total number of registered vehicles. Monthly values are interprelated from annual values (millions).
NCR	New Car Registrations
NTR	New Truck Registrations
NTR%	NTR as a percentage of total new vehicle registrations (NCR + NTR).
FLEETR	Registrations of fleets of 10 or more vehicles.
FCR	New Foreign Car Registrations

OTHER

CPI	"Consumer Price Index for All Workers" 1967 = 100.
HUDI	HUD interest contract rates on new commitments for conventional first mortgages.
HMI	Interest on conventional first mortgages on new home purchases.
LS	Liquor sales as the sum of "liquor store" sales and "drinking places" sales (millions of 72$).

TABLE 4.2 EMPIRICAL RESULTS OF THE MODIFIED YEATS PROCEDURE

(One Month Lag Unless Otherwise Indicated)

VARIABLE	t-STATISTIC[a]	COEF.
Driving Activity:		
VMT	3.62***	1.062
GSALES	3.20***	0.655
GMILES	3.23***	0.655
STATSLS	-0.68	-0.178
MPG	3.02***	0.49
Driving Cost:		
GAS$	-1.51*	-0.399
CAR COST	-1.50*	-0.929
Income and Employment:		
UNEMP	1.52*	0.224
UNEMP%	1.30*	0.196
YUNEMP	1.00	0.149
DUNEMP	0.85	0.112
EMP	0.90	1.106
EMP%	-0.22	-0.345
DI [b]	0.45	0.374
DIPC [b]	-0.05	-0.035
PI [b]	0.67	0.709
EARN	1.63*	1.296
RS	0.90	0.186
FRB	1.42*	0.528
C.IND [b]	0.34	0.182
L.IND [b,c]	1.36*	0.500
Demographic (Population and Motor Fleet):		
POP	1.49*	3.323
LF	2.31***	3.309
DRIVERS	0.31	1.618
CARREG	1.05	7.007
NCR [d]	-1.54*	-0.111
NTR	1.04	0.075
NTR % [d]	1.71**	0.195
FLEETR	-0.64	-0.025
FCR	-0.18	-0.010

TABLE 4.2 EMPIRICAL RESULTS OF THE MODIFIED YEATS PROCEDURE

(One Month Lag Unless Otherwise Indicated)

VARIABLE	t-STATISTIC[a]	COEF.
Other:		
CPI	-0.46	-0.617
HUDI	-1.93**	-0.259
HMI	0.88	0.212
LS	0.39	0.077

[a] t-Statistic values equal to, or greater than 1.30, 1.67, 2.00 are significant at the 80, 90, 95 percent level respectively.

*** = 95% level
** = 90% level
* = 80% level

[b] variables were seasonally adjusted before testing because seasonally unadjusted values are not available.

[c] t-Statistic is for a variable with a four month lag.

[d] t-Statistic is for a variable with a two month lag.

In the Demographic category, the labor force variable (LF) is the most significant at the 95 percent level. This variable may reflect driving age population. Population (POP) has the expected sign and is significant at the 80 percent level. New car registrations (NCR) reflect changes in the age distribution of the fleet. Since new cars tend to be safer than old cars, this variable probably reflects a change in the distribution of driving. The variable is significant at the 80 percent level and has the sign expected from the above hypothesis. New truck registration as a percent of total new vehicle registration has a 90 percent level of significance and has a positive sign which is consistent with the interpretation that it reflects changes in the heavy truck portion of the fleet. However, since this variable is defined as new truck registration over new car plus new truck registration, it may be dominated by the new car registrations in the denominator.

In the other category, an interest rate variable (HUDI) is significant at the 90 percent level. Liquor sales lagged one month was not significant.

Table 4.3 summarizes the variables found to be significant at the 80 percent level or higher. The table also identifies the lag found strongest and shows the prior studies reviewed in Section 2.1 which also found the variable significant.

4.2 Results of Autoregressive Modeling

As mentioned in Section 3, the purpose of this modeling is to test whether a variable adds significant explanatory power in predicting the highway fatality series given that prior knowledge about fatalities has been accounted for in a model. Prior knowledge is represented by a structure of lagged values of the fatality series. In addition, the autoregressive modeling procedure not only tests for significance of the variable, but determines which combination of past values of both the fatality series and the variable being tested is optimal in reducing the prediction error resultant from a specification. That is, if adding more terms of either the fatality or variable series will increase the prediction power of the model, adjusting properly for degrees of freedom, then the lagged terms of both series should be added until this prediction power (as measured by FPE) no longer is improved. It should be noted that low values of the FPE are improvements over higher values as the FPE is based on a minimization of differences between actual and historical values (again, properly adjusted for degrees of freedom).

While the FPE results are interesting by themselves in showing the optimal lag structure, and in making comparisons with optimal lag structures of other variables, the clear purpose of this modeling is to identify the lag structure for a particular variable which can be used in estimating a forecasting model which can be compared to historical data to make assessments about the variable's forecasting performance. This forecasting is done in the next section.

Table 4.3 Socio-Economic Influences Found Significant at the 80 Percent Level

Variable	t-statistic	lag (months)	Significance in Prior Studies
VMT	3.62***	1	Abt
GSALES	3.20***	1	
GMILES	3.23***	1	
GAS$	-1.51*	1	Abt
CARCOST	-1.50*	1	Land & McMilien; Peltzman
UNEMP	1.52*	1	Land & McMilien
EARN	1.63*	1	
FRB	1.42*	1	Abt; Joksch
L.IND	1.36*	4	
POP	1.49*	1	
LF	2.31***	1	
NCR	-1.54*	2	
NTR%	1.71**	2	
HUDI	-1.93**	1	

*** = Significant at the 95% level.
** = Significant at the 90% level.
* = Significant at the 80% level.

Of the 14 variables found significant at the 80 percent level by the modified Yeats procedure, ten have been carried to the autoregressive modeling. Population (POP) was eliminated because labor force (LF) appeared to be a substantially stronger variable which measured similar trends in population. Both new car registrations (NCR) and new truck registrations as a percentage of total new registrations (NTR%) are eliminated because they deal with driving distribution. (Note that NTR% is used with VMT in a multi-variable model described in Section 5.3.) Finally, CARCOST was eliminated because it is so similar to the real price of gasoline (GAS$).

As shown in Table 4.4, the FPE is given for just the lagged Y specification, which represents an autoregression model and will be referred to as such. The autoregressive model is compared with other models which may or may not have a lower prediction error. If a model does have a lower prediction error, and if the t-statistic values of some of its lagged terms are significant then it is judged to be an improvement over the autoregressive model. Such an improvement is evidence of statistical causality.

Referring again to Table 4.4, it is observed that Driving Activity variables—VMT, GMILES, GSALES—are all successful in making improvements over the autoregressive model, findings which do not disprove statistical causation at the point of optimal lag (lag = 1). The Driving Cost variable, GAS$, also has an optimal lag at one month, but is not as strong as the driving activity variables, probably because it is only one determinant of Driving Activity. The evidence does not support statistical causality of GAS$.

In the Income and Employment category, FRB and EMP do better than UNEMP%, EARN, and L.IND. The last variable has an optimal lag at three months. The labor force variable (LF) in the Demographic category has the lowest FPE of all the 11variables tested and, as mentioned before, may serve as a proxy for driving-age population. Finally, the interest rate variable, HUDI, at a two month lag does relatively well.

A caution must be noted when comparing the FPE's of Table 4.4. While a relatively low value at the optimal lag is a low sum of squared residual, adjusted for proper degrees of freedom, it is not necessarily true that lower FPE's are statistically different from higher values. Rather, such a comparison is a method for making judgments as to which variables to select for testing for ultimate inclusion in a structural regression model.

Table 4.4 Results of Autoregressive Modeling

Variables	FPE* (at Best Lag)	Best (Lag)	Statistical Causation
Autoregressive	.002352	1-4	
VMT	.002181	1	Yes
GSALES	.002167	1	Yes
GMILES	.002160	1	Yes
GAS$	.002366	1	No
UNEMP %	.002311	1-2	Yes
EMP	.002289	1	Yes
EARN	.002333	1	Yes
FRB	.002280	1	Yes
L.IND	.002385	1-3	No
LF	.002140	1	Yes
HUDI	.002207	1	Yes

*Final Prediction Error.

4.3 Forecasting Performance

Eight variables were selected from the autoregressive modeling procedure: VMT, FRB, GSALES, GMILES, LF, HUDI, EARN, and L.IND. A ninth variable--DUNEMP--was added to test the power of the Yeats procedure in screening variables. DUNEMP, the duration of unemployment, was not significant at any lag period in the Yeats procedure and the expectation was that it should not do better in forecasting than variables that did well.

There is, of course, always the possibility that a variable that does well in the sample estimation period will not forecast well outside the sample period for a number of reasons. This possibility, however, is somewhat lessened by the fact that the variables used in forecasting were selected partly on the basis that each did well over the entire fatality series, January 1975 to December 1982. It is the last 24 months of that series which the model will be tested on.

The methodology for forecasting is:

1. Estimate models using the optimal lags determined in the autoregressive modeling procedure for each of the nine variables for the period ending in December 1980.

2. Use the parameters of these models to forecast fatalities for the next 24 months (January 1981 to December 1982).

3. Compare the forecast results with the historical values and evaluate the performance.

In making the forecasts from the models which all contain lagged fatality series values (lagged endogenous variables), historical values were used and not forecast

* This amounts to making a series of one-month forecasts.

values of fatalities for these lagged terms*. Thus an under or over estimation of fatalities is not carried over into the next period—errors are not cumulative. This, of course, is not possible in many real forecasting situations. For purposes of evaluating the strengths of forecasting models, however, it is possible and the technique is known as unconditional, ex-post forecasting.

Table 4.5 presents the percentage error of the fatality estimate for the 24-month period, the average absolute value of the percent error over the full 24 months, and the average absolute percent error for each of the four six-month periods. As well as giving forecasting results for nine variables, the results of the autoregressive model are given and it is expected that the other models would noticeably improve the forecasts to be of value. The column labelled S.E. is the standard error of the fatality count for the month as a percent of the count assuming a Poisson process, and it is given to aid in the interpretation of the significance of the other percent error columns.

As judged by the overall average absolute percent error, adding VMT, FRB, HUDI or EARN to the autoregressive model causes improvement. In the first six-month period, it is interesting to observe that none of the models perform as well as they do in the next three periods. This result is influenced by the large underestimate* of the January 1981 historical value for all models except the one containing the LF. All the models then proceed to do better than in the first six months, with all models making their best forecasts is the last six-month period.

The average absolute percent errors range from a low of 3.02 for the EARN variable to a high of 3.66 for LF with the autoregressive model having a 3.33 value. The question that naturally arises, is: Are these differences from the autoregressive model statistically significant? A paired t-test was used to find that none of the models were significantly different from the autoregressive model at the 95 percent level.

* A minus sign indicates an underestimate.

TABLE 4.5 FORECASTING PERFORMANCE OF VARIABLES USE IN THE AUTOREGRESSIVE MODELING PROCEDURE FOR DETERMINING OPTIMAL LAG STRUCTURE

Percent Error (Estimate-Actual/Actual)

YEAR	MO.	S.E.	AUTO	VMT	FRB	GSALES	GMILES	LF	HUDI	EARN	LIND	DUNEMP
1981	1	1.69	-9.33	-8.98	-9.34	-9.17	-9.39	1.06	-10.63	-9.68	-10.96	-9.91
1981	2	1.72	-0.91	0.82	-1.03	-0.48	-0.53	1.29	0.33	0.17	-0.26	-1.69
1981	3	1.65	7.13	4.61	6.36	2.69	2.52	6.32	9.16	6.22	7.56	7.03
1981	4	1.58	-1.99	-0.71	-2.29	-1.44	-1.42	0.09	0.66	-1.43	0.31	-4.35
1981	5	1.56	9.05	8.42	9.03	9.02	9.03	9.36	5.86	9.58	8.36	7.09
1981	6	1.51	3.7	2.74	4.51	4.36	4.39	6.24	-0.22	3.23	0.64	5.22
1981	7	1.43	-0.78	-1.7	-0.33	0.53	0.56	-8.64	-0.43	-1.43	1.03	-0.23
1981	8	1.44	1.69	-0.19	2.93	2.42	2.42	1.48	-0.49	1.38	1.93	2.65
1981	9	1.53	4.56	4.55	3.93	1.42	1.29	4.93	3.21	4.7	4.69	5.4
1981	10	1.55	4.65	7.34	2.35	7.16	7.22	2.96	4.19	3.39	4.11	3.66
1981	11	1.58	-2.21	-1.8	-3.65	-1.92	-1.91	-0.38	0.56	-1.62	0.57	-4.02
1981	12	1.58	3.56	2.74	2.05	1.26	1.17	3.22	6.23	4.01	4.29	4.13
1982	1	1.89	4.73	4.82	2.44	6.64	6.63	1.98	1.62	3.91	2.93	5.65
1982	2	1.9	2.32	-0.41	0.78	-2.25	-2.3	3.2	1.56	1.81	3.34	3.57
1982	3	1.7	-2.55	-0.75	-2.07	1.18	1.38	-2.42	-0.79	-0.49	-1.87	0.01
1982	4	1.67	1.95	0.57	1.	2.34	2.42	2.07	3.58	1.66	2.05	1.64
1982	5	1.62	7.04	7.28	6.02	6.94	7.01	6.89	5.93	7.61	8.07	4.62
1982	6	1.6	3.39	4.54	2.96	2.55	2.61	8.15	3.29	3.49	1.97	4.33
1982	7	1.54	0.62	-0.8	0.62	1.66	1.78	-5.94	-0.93	-0.65	0.5	1.66
1982	8	1.53	1.93	1.37	3.05	2.25	2.34	2.75	2.11	2.12	2.68	0.37
1982	9	1.59	-0.2	-0.79	-0.53	-3.6	-3.65	-0.02	3.82	-0.1	-2.03	-2.83
1982	10	1.58	0.88	2.56	-1.13	4.24	4.39	1.92	0.35	0.03	2.56	1.2
1982	11	1.64	-1.99	-1.38	-3.44	-3.63	-3.63	-4.54	3.04	-1.93	-2.52	-1.14
1982	12	1.6	-2.71	-2.81	-3.51	-1.97	-1.89	-2.09	-3.94	-1.94	-2.9	-3.6

AVERAGE ABSOLUTE PERCENT ERROR

	PERIOD	S.E.	AUTO	VMT	FRB	GSALES	GMILES	LF	HUDI	EARN	LIND	DUNEMP
	81-1 Thru 82-12	1.61	3.33	3.03	3.14	3.38	3.41	3.66	3.04	3.02	3.26	3.58
1	81-1 Thru 81-6	1.62	5.35	4.38	5.43	4.53	4.55	4.06	4.48	5.05	4.68	5.88
2	81-7 Thru 81-12	1.52	2.91	3.05	2.54	2.45	2.43	3.60	2.52	2.76	2.77	3.35
3	82-1 Thru 82-6	1.73	3.66	3.06	2.55	3.65	3.73	4.12	2.80	3.16	3.37	3.30
4	82-7 Thru 82-12	1.58	1.39	1.62	2.05	2.89	2.95	2.88	2.37	1.13	2.20	1.80

5.0 Multiple Variable Models of Highway Fatalities

In the previous section, the relationship between each of 34 socio-economic variables and highway fatalities was evaluated based on the variable's ability to forecast highway fatalities. In this section, more intuitive models of how these variables influence highway fatalities are tested. In particular, the models of fatalities are not limited to models which use past information to predict fatalities in the future. Since fatalities are a by-product of driving, fatalities and driving (VMT) are contemporaneous. This section describes a preliminary attempt to build multiple variable models of highway fatalities. Section 6.0 then uses these models to better understand the forces causing fatalities to drop by 10 percent in 1982.

The approach to building multiple variable models of fatalities is to: 1) Develop and test models with socio-economic variables hypothesized to influence fatalities through their influence on total driving (VMT); 2) Further test the hypothesis that the path of influence on fatalities is through VMT by using "path analysis"; and 3) Develop fatality models which include variables hypothesized to influence the amount of risky driving by including these variables with VMT. The best of the models developed in steps 1) and 3) can then be combined into models which describe the influence of socio-economic factors on highway fatalities. Sections 5.1, 5.2, and 5.3 describe these three steps.

5.1 Economic Influences Acting Through Total Driving

This section describes four models of highway fatalities which are based on economic variables hypothesized to influence total driving. Table 5.1 presents the four models. These models are identical except for the measure used to represent the level of consumer income or spending. The measures of consumer income or spending are: retail sales (RS); average weekly earnings per production worker (EARN); and total personal income (PI). The FRB production index is included in this analysis because it has been widely used to measure the performance of the economy. It is probably closely related to income and spending. The level of consumer income influences fatalities by increasing

Table 5.1 Models of Influence Acting Through Driving Activity

1) $\Delta \log \hat{F}_t = a + (\text{seasonal dummies}) + b\, \Delta \log (PI_t) + c\, \Delta \log (RCOST_t) +$
$RH01\, (\Delta \log F_{t-1} - \Delta \log \hat{\hat{F}}_{t-1})$

2) $\Delta \log \hat{F}_t = a + (\text{seasonal dummies}) + b\, \Delta \log (FRB_{t-1}) + c\, \Delta \log (RCOST_t) +$
$RH01\, (\Delta \log F_{t-1} - \Delta \log \hat{\hat{F}}_{t-1})$

3) $\Delta \log \hat{F}_t = a + (\text{seasonal dummies}) + b\, \Delta \log (RS_t) + c\, \Delta \log (RCOST_t) +$
$RH01\, (\Delta \log F_{t-1} - \Delta \log \hat{\hat{F}}_{t-1})$

4) $\Delta \log \hat{F}_t = a + (\text{seasonal dummies}) + b\, \Delta \log (EARN_t) + c\, \Delta \log (RCOST_{(t)}) +$
$RH01\, (\Delta \log F_{t-1} - \Delta \log \hat{\hat{F}}_{t-1})$

where:

F_t = fatalities in time t (actual)

$\hat{F}_t$= estimated fatalities in time period t.

$\hat{\hat{F}}_{t-1}$= estimated fatalities in period t-1 without the RH01 term (the correction for first order autocorrelation).

PI_t = total personal income.

$RCOST_t$ = the relative cost of driving; index of private transport cost (CARCOST) divided by the consumer price index (CPI).

FRB_t = Federal Reserve Board production index.

RS_t = retail sales.

$EARN_t$ = average weekly earnings per production worker.

a, b, c = estimated coefficients.

RH01 = first order autocorrelation correction factor.

driving activity. Consumers spend more when they have more income and part of that spending goes into driving. The fraction of income they spend on driving depends on the cost of driving in relation to the cost of other goods and services. If driving becomes relatively more expensive, consumers will spend less on driving by accumulating trips, ride sharing, using public transit, and eliminating trips. The RCOST term in the models is included to capture this effect. Note that for the RS_t, PI_t and $EARN_t$ variables no lags were found to be significant at the 95 percent level in the modified Yeats procedure. Results for these and other variables are presented in Appendix B. FRB_{t-1} (with a one-month lag), although not found to be significant at the 90 percent level by the modified Yeats procedure, appears to be the best lag structure for this variable. Finally, the relative cost variable (RCOST) was used with no lag (t) because we hypothesized current costs to influence current driving decisions.

Table 5.2 presents the results of fitting these equations to the full set (1975-1982) of monthly fatality data. Also included in Table 5.2 is the result of fitting the $VMT_{(t)}$ variable for comparison with the other equations. All equations were fit using GLS with a first-order autocorrelation correction (RHO1). The coefficients of all the income/spending variables have the anticipated sign and all but FRB_{t-1} are significant at the 95 percent level. The $RCOST_t$ term has the anticipated sign but is not significant at the 90 percent level except in the FRB model. The equations all perform well and the standard error of estimate of the best differs from the worst by only 10 percent.

Table 5.3 presents the results of fitting these models to a part of the data for use in the forecasting test. As shown in this table, the $RCOST_t$ variable has much larger coefficients and higher significance. This instability along with the poor t-statistics in Table 5.2 make the $RCOST_t$ term a liability in the forecasting test.* Table also 5.3 shows the result of refitting these models without the RCOST term.

*A possible explanation for this instability is that while we have been using RCOST, which has the gas price included as a major component, to measure the attractiveness of driving the amount of driving also influences the demand for gas and this influences its price. This explanation is consistent with falling gas prices in 1982 and the smaller negative coefficients when the 1981 and 1982 data are included.

Table 5.2 Results of Estimating the Driving Activity Equations (1975 thru 1982)

Equation	Coefficient and t-statistic b	c(RCOST)	Adjusted R^2	Standard error of estimate	RHO1*
(0) $VMT_{(t)}$	1.413 (4.96)	—	.8916	.0421	-.246
(1) $PI_{(t)}$	2.067 (1.98)	-1.055 (-1.29)	.8684	.0465	-.249
(2) $FRB_{(t-1)}$	.621 (1.71)	-1.658 (-2.01)	.8706	.0458	-.205
(3) $RS_{(t)}$	.848 (4.38)	-1.130 (-1.57)	.8907	.0430	-.303
(4) $EARN_{(t)}$	1.667 (2.12)	-.990 (-1.18)	.8668	.0462	-.201

*RHO1 is the correction factor for first-order autocorrelaton.

Table 5.3 Results of Estimating the Driving Activity Equations (1975 thru 1980)

Equation	Coefficient and t-statistic b	c(RCOST)	Adjusted R^2	Standard error of estimate	RHO1*
(0) $VMT_{(t)}$	1.426 (4.37)	—	.8924	.0435	-.261
(1) $PI_{(t)}$	1.905 (1.54)	-2.113 (-1.94)	.8731	.0476	-.293
	2.311 (1.86)	—	.8656	.0487	-.271
(2) $FRB_{(t-1)}$	.632 (1.40)	-2.927 (-2.72)	.8804	.0460	-.258
	.432 (.90)	—	.8644	.0485	-.213
(3) $RS_{(t)}$	.772 (3.61)	-2.304 (-2.40)	.8943	.0440	-.338
	.770 (3.47)	—	.8831	.0458	-.299
(4) $EARN_{(t)}$	1.183 (1.14)	-2.159 (1.88)	.8673	.0480	-.233
	1.618 (1.56)	—	.8597	.0490	-.200

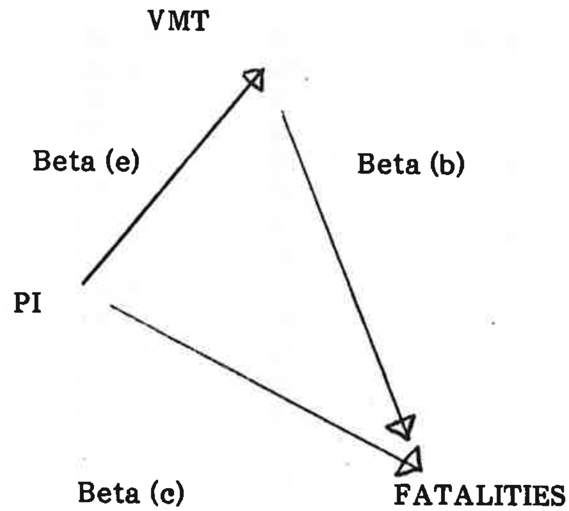
Table 5.4 Forecasting Performance of the Driving Activity Models (Percent Error (Est.-Actual/Actual))

	S.E.	VMT	EARN	RS	FRB	PI
1981 1	1.69	-7.4	-9.29	-9.09	-11.82	10.47
1981 2	1.72	-0.92	1.91	2.21	5.75	4.98
1981 3	1.65	10.28	9.85	8.57	8.73	8.35
1981 4	1.58	-5.36	-3.94	-4.42	-4.68	-5.19
1981 5	1.56	8.38	9.9	8.67	11.03	10.80
1981 6	1.51	-3.56	-4.63	-3.79	-3.67	-4.32
1981 7	1.43	-5.30	-3.52	-2.68	-3.53	-3.22
1981 8	1.44	2.84	2.31	1.07	2.93	2.80
1981 9	1.53	5.94	0.6	8.39	2.17	1.82
1981 10	1.55	3.05	5.5	2.69	2.86	2.84
1981 11	1.58	-5.14	-4.27	-6.99	-5.65	-5.96
1981 12	1.58	3.25	1.91	5.83	2.63	2.16
1982 1	1.89	-2.45	1.21	-.45	0.36	0.29
1982 2	1.9	3.37	2.55	-.53	-0.21	0.13
1982 3	1.7	-5.29	-4.57	-4.05	-3.41	-4.46
1982 4	1.67	2.35	2.49	3.35	1.37	3.01
1982 5	1.62	7.35	6.79	5.06	6.46	7.74
1982 6	1.6	-2.03	-1.55	-1.69	0.29	-2.61
1982 7	1.54	-1.14	-0.64	1.11	-1.56	-2.04
1982 8	1.53	-0.46	-0.06	-3.47	0.39	-0.84
1982 9	1.59	1.04	-3.21	1.88	-2.29	-2.78
1982 10	1.58	1.72	2.19	0.01	0.79	1.03
1982 11	1.64	-2.08	-0.67	0.36	-2.55	-0.97
1982 12	1.6	-1.74	-0.03	1.19	1.84	-2.05

Average Absolute Percent Error

	S.E.	VMT	EARN	RS	FRB	PI
PERIOD 81-1 82-12	1.61	3.85	3.48	3.44	3.62	3.79
81-1 81-6	1.62	5.98	6.59	6.12	7.61	7.35
81-7 81-12	1.52	4.25	3.02	3.78	3.30	3.13
82-1 82-6	1.73	3.81	3.19	2.52	2.02	3.04
82-7 82-12	1.58	1.36	1.13	1.34	1.57	1.62

Figure 5-1 Path Analysis



Equations:

1. $\Delta \log F_t = a + b\Delta\log VMT_t + c\Delta\log PI_t$ + seasonal dummies

2. $\Delta \log VMT_t = d + e\Delta\log PI_t$ + seasonal dummies

Direct Effect = Beta (c)
Indirect Effect = Beta (e) x Beta (b)

Total Effect = Direct + Indirect

After deleting the $RCOST_t$ term, the coefficients for all of the income/spending variables except FRB are similar to those estimated for the full 1975-1982 period. The FRB_{t-1} variable is not significant even at the 80 percent level. It is included in further analysis because other researchers have found it to be useful in explaining highway fatalities.

Table 5.4 presents the results of using the equations in Table 5.3 to forecast the period 1981 through 1982. In these forecasts actual historic values are always used for the independent variables and the last period residual (t-1) is used with RHO1 to adjust the current estimate of fatalities. Table 5.4 presents the percentage error of the fatality estimate by month, the average absolute value of the percent error over the full 24 months, and the average absolute percent error for each of the four six-month periods in the sample.

The column labeled S.E. is the standard error of the fatality count for the month as a percent of the count assuming a Poisson process. This column is meant to help interpret the significance of the other percent error columns.

In general these models perform well, estimate fatalities quite accurately, and contain no systematic under-or over-estimation. Over the full 24 month period the models using economic variables appear to perform as well as the model using the VMT. During the 12 months in the middle of the period, 81-7 through 82-6, the economic variable models appear to perform slightly better than the VMT model.

5.2 Path Analysis

The hypothesis behind these economic models of highway fatalities is that they influence VMT which influences fatalities. To test this path of influence, a number of regression equations were estimated. Figure 5-1 presents a summary of this path analysis method which is described in Section 3.3. The heart of the method is the use of Beta coefficients which are standardized coefficients. A Beta coefficient of .5 means that a standard deviation change in the independent variable leads to a .5 standard deviation change in the dependent variable. Thus, the size of the Beta coefficient can be used to compare the size of two variable effects. The Beta coefficients from the two equations shown in Figure 5-1 permit the total effect of a variable to be divided into an indirect effect, which is the variable's influence on VMT and subsequently on fatalities, and a direct effect on fatalities.

Table 5.5 presents Beta b, c, and e along with the t-statistics for the economic variables used in the models defined in Figure 5.1. The corrected R^2 for the fatality (R^2_F) and VMT (R^2_V) equations are also shown in Table 5.5. While the direct effects of the variables are not significant (except RS_t), the indirect effects are not significant either. When neither effect is significantly different from zero, no statistical interpretation of the path of influence is possible. However, the very strong Beta(e)'s between the economic variables (except FRB) and VMT and the high R^2_V's indicate the substantial influence of these variables on VMT. So, the evidence in Table 5.5 suggests that the economic variables have a substantial influence on VMT and we interpret these economic variables as surrogates for VMT in models of highway fatalities.

Since Beta(c) is significant, retail sales (RS_t) has a direct effect on fatalities beyond its effect on VMT. We interpret this to mean that RS_t, a measure of consumer spending, somehow influences the distribution of total driving. Driving in riskier than average situations increases with RS. In the next subsection, variables which we hypothesize to influence risky driving are tested.

5.3 Influences on Fatalities via the Distribution of Driving

In this section, models which explain fatalities in terms of changes in the distribution of VMT are sought. The approach is to include terms which might indicate changes in the fraction of total driving which is in risky situations: teenage, rural, vacation, late night, weekend, DWI, or heavy truck driving. These terms were added to a model with VMT alone and the results compared to the VMT model.

Table 5-5. Path Analysis Results

Variable	(Direct Effect) Beta (c)	Beta (b)	Beta (e)	Indirect Effect Beta (b) x Beta (e)	$\bar{R}^2_F$	$\bar{R}^2_V$
PI_t	.041 (1.09)*	.739 (4.45)	.064 (2.67)	.047 (1.25)	.893	.949
FRB_{t-1}	.076 (1.04)	.728 (4.39)	.049 (.95)	.036 (.49)	.892	.948
RS_t	.494 (2.66)	.591 (3.46)	.452 (4.51)	.267 (1.44)	.902	.955
EARN (t)	.043 (.77)	.737 (4.24)	.122 (3.76)	.090 (1.61)	.891	.952

*t-statistics are in parentheses

Table 5.6 presents five socio-economic variables and hypotheses describing how each may influence the distribution of driving or indicate changes in the distribution of driving. To test the hypothesis that the variable has an influence on the distribution of driving, we estimated the coefficients of an equation which explains fatalities using the seasonal dummies, VMT and the variable:

$$\Delta \log F_t = a + b \Delta \log VMT_t + c \Delta \log (variable_t) + \text{seasonal dummies}$$

The effect of total driving is captured by the VMT_t term, so the variable captures the effect of any change in the distribution of driving.

The percentage of new vehicle registrations which are trucks (percentNTR) is treated slightly differently. The %NTR represents a change in the fleet composition itself and the motor vehicle fleet should be more indicative of total fatalities than the new registrations. So, %NTR was used directly, with the change in VMT to estimate fatalities:

$$\Delta \log F_t = a + b\Delta \log VMT_t + c \log \%NTR_t + \text{seasonal dummies}$$

Table 5-7 presents the results estimating these equations using GLS and correcting for first-order autocorrelation. Table 5-7 also presents the results of estimating an equation with VMT alone.

None of the coefficients is significant at the 95 percent confidence level but all of the coefficients have the expected sign except EMP_t. The level of employment is closely related to other indicators of economic activity which tend to positively influence VMT. The EMP_t coefficient is dominated by this effect rather than its effect on commuting travel.

The data in Appendix B show a high t-statistic for liquor sales with no lag. The insignificant coefficient shown in Table 5.7 suggests that the high t-statistic is attributable to its behavior as an indicator of retail sales which effects total driving. As a measure of drunk driving, it does not perform well.

Table 5.6

Influences on the Risk Distribution of Driving

Variable	Hypothesized Effect
Teenage Unemployment (YUN)	Higher teenage unemployment causes lower teenage driving and lower fataliity rates.
Liquor Sales (LS)	Higher liquor sales causes more drunk driving and higher fatality levels.
National Park Visits (NPM)	Higher National Park visits indicate more vacation travel and higher fatality levels.
Employment (EMP)	Higher employment causes more commuter travel and lower fatality levels.
%New registrations which are trucks (%NTR)	Higher proportion of trucks in the motor vehicle fleet indicates a higher proportion of heavy truck VMT and higher fatality levels.

Table 5-7 Variables Influencing the Distribution of Driving

Equation	Coefficient (t-statistic) VMT_t	Driving Distribution Variable	$\bar{R}^2$	Standard Error
VMT alone:	1.413 (4.96)	—	.8916	.0421
VMT plus:				
$YUN.M_t$	1.343 (4.52)	-.030 (-.23)	.8912	.0424
$\%NTR_t$	1.392 (4.95)	.068 (1.72)	.8952	.0417
LS_t	1.282 (9.29)	.252 (1.36)	.8919	.0419
NPM_t	1.489 (5.03)	-.043 (-1.01)	.8906	.0421
EMP_t	1.353 (4.62)	.967 (.88)	.8917	.0422

5.4 Summary

In this section we have identified four economic variables which we hypothesize to influence fatalities by influencing total driving. These variables are personal income (PI), retail sales (RS), average weekly earnings of production workers (EARN) and the FRB production index (FRB). With the exception of FRB, the coefficients of these variables were found to be significant at the 95 percent confidence level. Tests of the monthly forecasting performance of these models found them all to be roughly comparable to the VMT model.

Using the path analysis technique, we established that these variables do not influence fatalities solely through VMT, though only retail sales has a statistically significant direct effect (not through VMT) on fatalities.

Finally, none of the five socio-economic variables hypothesized to influence the risk distribution of driving was found to add significantly to a model of fatalities based on VMT. This may be because there are many conflicting forces acting on aggregate fatality levels. The result suggests that examination of socio-economic influences on high-risk strata of fatalities might be more productive.

A synthesis of models describing the amount and risk distribution of driving is precluded by the lack of a significant model of influences on high-risk driving. As a result, the best models of highway fatalities are those which include the seasonal variables and either VMT or one of the four economic variables hypothesized to influence fatalities through VMT. In the next section, these five models will be used to better understand the 10 percent drop in fatalities which occurred in 1982.

6.0 The 1982 Decline in Highway Fatalities

The 10 percent fatality decline in 1982 is peculiar because VMT actually increased by one percent in 1982. Thus, the highway fatality rate (per vehicle mile) is lower than it has been for many years. In this section, the 10 percent decline will be examined by using the models developed in Section 5.0 to calculate the level of fatalities which would be expected based on the actual levels of VMT, or its surrogates: personal income (PI), retail sales (RS), earnings (EARN) and the FRB production index (FRB). The question we hope to answer with this examination is "Does 1982 represent a significant departure from the relationship between fatalities and VMT, or its surrogates, which held prior to 1982?"

6.1 Model Performance in 1982

Table 6.1 presents the results of using the models developed in Section 5.0 (Table 5.3) to estimate highway fatalities in 1982. All of the models start with actual fatalities in December 1981, and estimate monthly changes in fatalities for the next twelve months based on the seasonal dummy variables and VMT or the socio-economic surrogates.* The model coefficients were obtained by estimating the model on monthly data between 1975 and 1980, so the model can be thought to represent the relationship between fatalities and VMT over this period. The complete equations for the five models used to produce these estimates and all others in this section are presented in Appendix C.

*Note that tests of forecasting performance in Sections 4.0 and 5.0 were sequences of one-month forecasts while this section presents one-year forecasts. That is, in Sections 4.0 and 5.0 the autocorrelation correction, RH01, was used with the past month's actual and estimated fatalities to improve the fatality estimates. In this section, only December fatalities are used. While the RH01 does have an effect on January, the effect is negligible in other months. The main purpose of the RH01 factor is to get better estimates of the coefficients. The form of the model (estimating percent change in fatalities) was chosen because of its statistical properties. Errors in this forcasting test accumulate because each month's forecast is the product of the prior month's fatalities times the estimated percentage change in fatalities, and since the level of the prior month's fatalities was estimated in the same way, the errors tend to build-up until January when the forecast is based on actual December values.

TABLE 6.1

FORECAST RESULTS FOR 1982

MONTH	ACTUAL*	VMT	PI	FRB	RS	EARN
		(Percent Error	(Estimate - Actual)/Actual)			
January	2809.53	-0.11	1.50	2.06	2.42	2.61
February	2782.00	2.43	1.55	1.77	1.94	5.35
March	3451.72	-2.26	-2.85	-1.65	-2.13	1.06
April	3589.96	-1.13	-1.06	-1.00	-0.03	2.77
May	3816.04	6.44	7.07	5.53	5.65	10.08
June	3907.48	6.35	6.60	7.28	5.61	9.90
July	4221.04	5.12	4.32	5.96	6.77	9.16
August	4255.19	4.32	2.84	6.10	3.41	8.95
September	3971.96	5.18	-.45	3.70	4.32	5.41
October	4002.04	7.21	-.28	4.03	4.60	7.03
November	3724.96	5.57	-1.19	1.45	5.06	6.67
December	3918.20	3.32	-3.46	-0.93	3.93	6.57
TOTAL	44,450	46,121	45,021	45,796	46,061	47,350
PERCENT ERROR		3.76	1.22	3.03	3.62	6.50
RMS PERCENT ERROR		4.65	3.51	4.08	4.24	6.92
ESTIMATED PERCENT CHANGE FROM 1981 ACTUAL	-9.70	-6.30	-8.53	-6.96	-6.42	-3.80

*Fatality counts are adjusted only for working day/trading day variation and not for seasonality. The fatality counts are based on values contained in the FARS data base as of February 1983.

As the last row in Table 6.1 shows, all of the models underestimate the size of the 1982 fatality decline. The model using personal income (PI) comes the closest, estimating an 8.5 percent decrease in fatalities. Average weekly earnings per production worker (EARN) does the worst, estimating only a 3.8 percent decrease. The VMT model is in the middle, estimating a 6.3 percent decrease. The root mean square (RMS)* error is between 3.5 and 4.7 percent for all models except the EARN model. All of the models substantially overestimated May and then gradually correct that overestimate.

While the RMS error and the estimated fatality reduction from 1981 look good for all the models (except EARN), 1982 might still be substantially different than other years. If the RMS error is substantially higher for 1982 than for other years, the relationship between fatalities and VMT or its surrogates may have changed. Table 6.2 presents the RMS and total percent error for the five models for each year between 1975 and 1982. Each of the yearly estimates was derived from forecasts starting with the prior year's December value inserted into the equations presented in Appendix C. Figures 6.1 through 6.5 present plots of the estimated (closed box) and actual (opened box) fatalities by month which were developed.

The RMS errors generally fall between 4 and 6 percent. The RMS errors for 1975 are higher than for other years, except for the FRB model where they are lower. The relationship between VMT and highway fatalities appears to be roughly the same in 1982 as in prior years back through 1976. The same conclusion is reached for each of the economic variables except EARN where the RMS error appears to be somewhat higher in 1982.

$$RMS = 1/T \sum_{t=1}^{T} ((E_t - A_t)/A_t)^2$$

Where E_t = Estimated value in time period t.

A_t = Actual value in time period t.

T = Number of forecasts

*The root mean square (RMS) error is often used to measure forecasting performance. Large individual errors are heavily penalized and positive and negative errors do not cancel. RMS is defined as:

TABLE 6.2

PERCENT ERRORS FOR ONE YEAR FORECASTS

Percent Forecast Error: (Estimated-Actual)/Actual

YEAR	ERROR MEASURE	VMT	PI	FRB	RS	EARN
1975	Ave.	7.3	8.4	1.4	8.9	8.2
	RMS	7.6	8.6	2.3	9.0	8.4
1976	Ave.	-3.2	-1.3	0.3	-0.5	-0.4
	RMS	4.7	3.6	3.9	3.3	4.0
1977	Ave.	-0.5	2.4	-3.5	-4.4	-2.6
	RMS	2.6	4.1	4.5	5.2	4.3
1978	Ave.	-2.3	-3.1	-4.0	-4.5	-3.0
	RMS	3.8	6.0	6.1	6.0	6.1
1979	Ave.	-0.2	-2.0	1.0	-1.4	-3.3
	RMS	4.7	4.7	4.3	4.2	5.6
1980	Ave.	0.2	-1.0	2.7	0.4	-0.7
	RMS	4.8	4.4	6.0	4.5	4.6
1981	Ave.	-0.3	0.3	1.7	-0.8	0.2
	RMS	5.2	5.5	6.4	4.9	5.3
1982	Ave.	3.8	1.2	3.0	3.6	6.5
	RMS	4.7	3.5	4.1	4.2	6.9
1975-1982	RMS	4.9	5.2	4.9	5.3	5.8
1976-1982	RMS	4.4	4.6	5.1	4.7	5.4

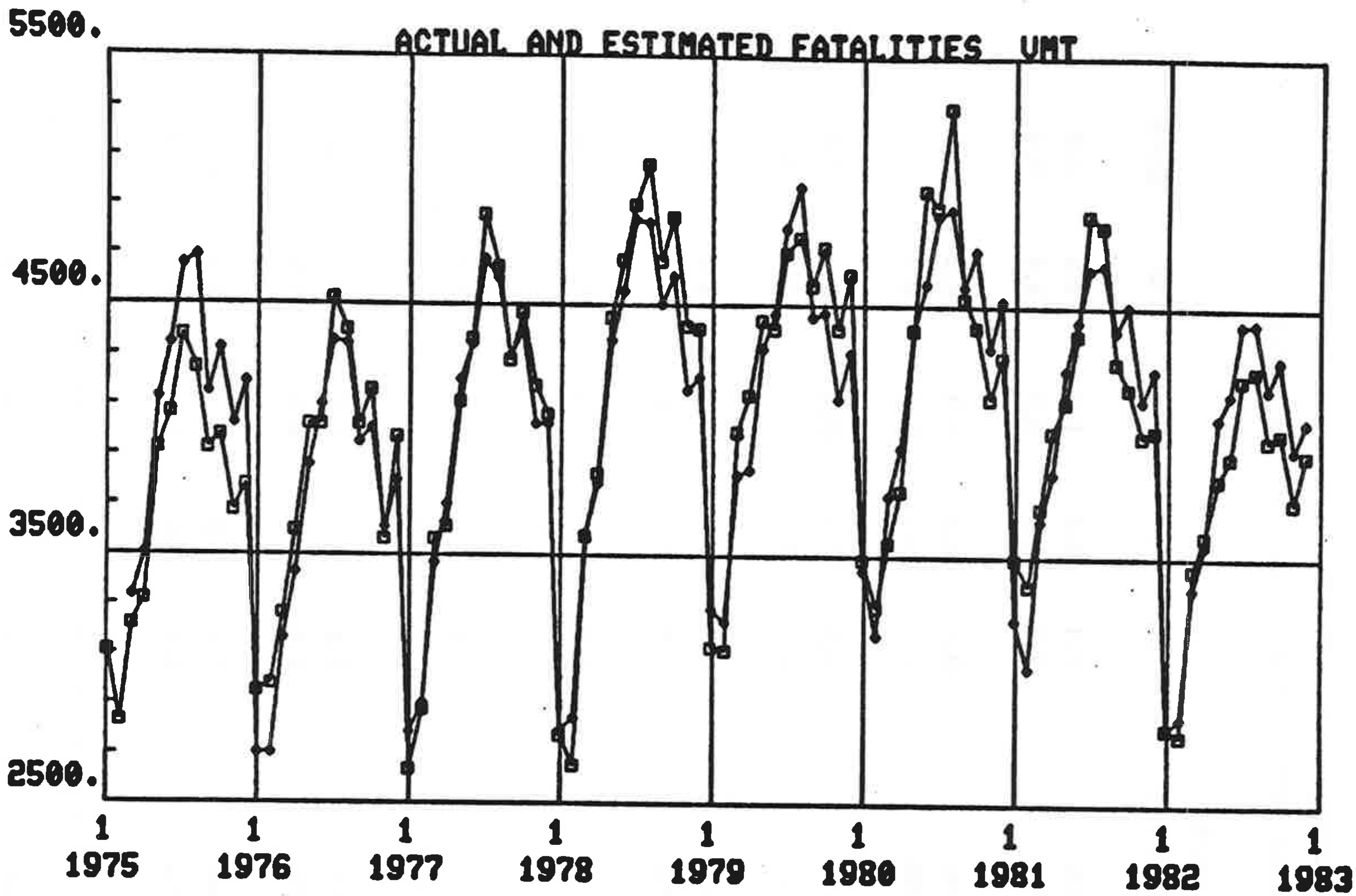


FIGURE 6.1: VMT Model Performance

Estimated (closed box)
Actual (opened box)

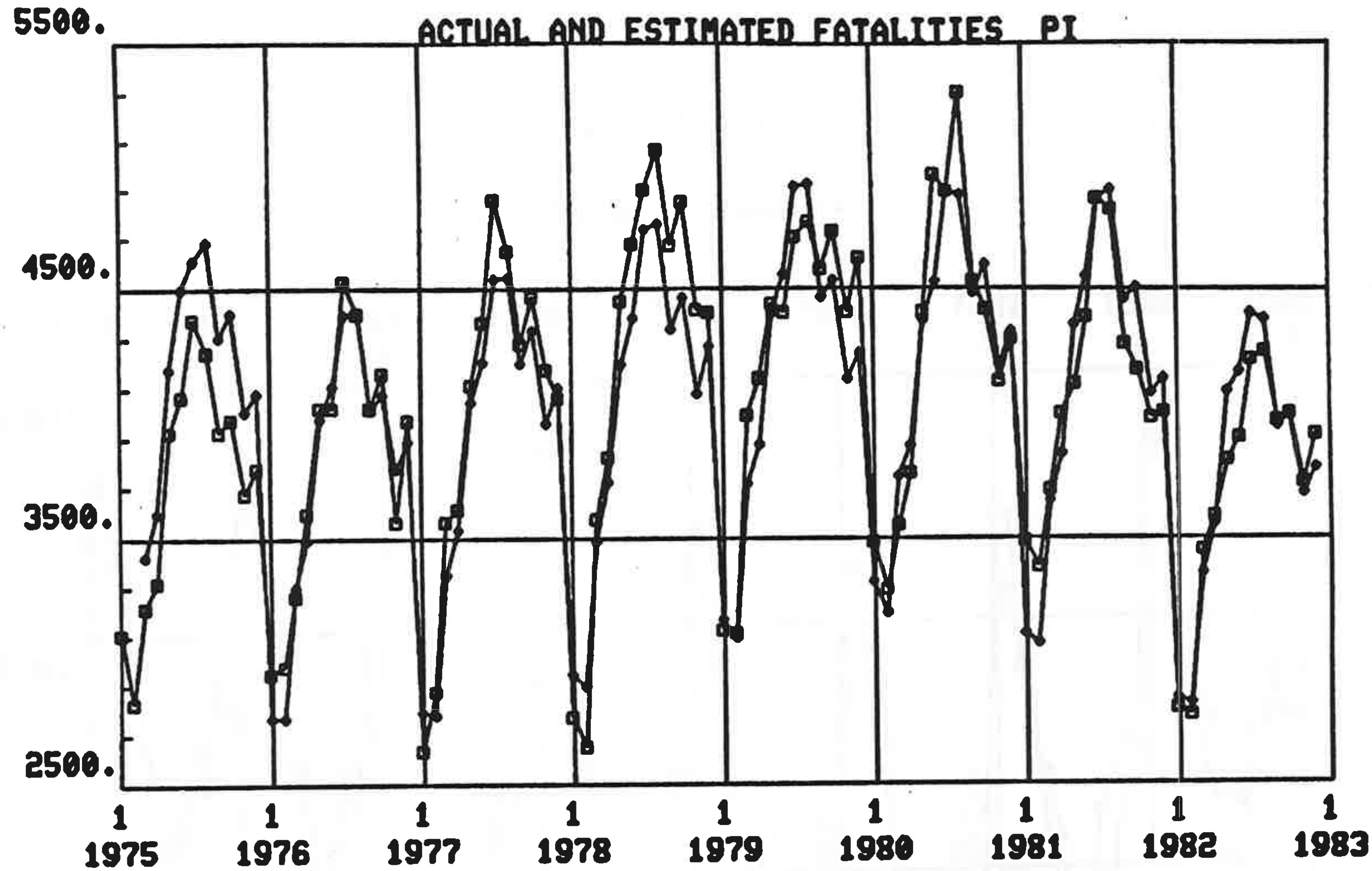


Figure 6.2: PI Model Performance

Estimated (closed box)
Actual (opened box)

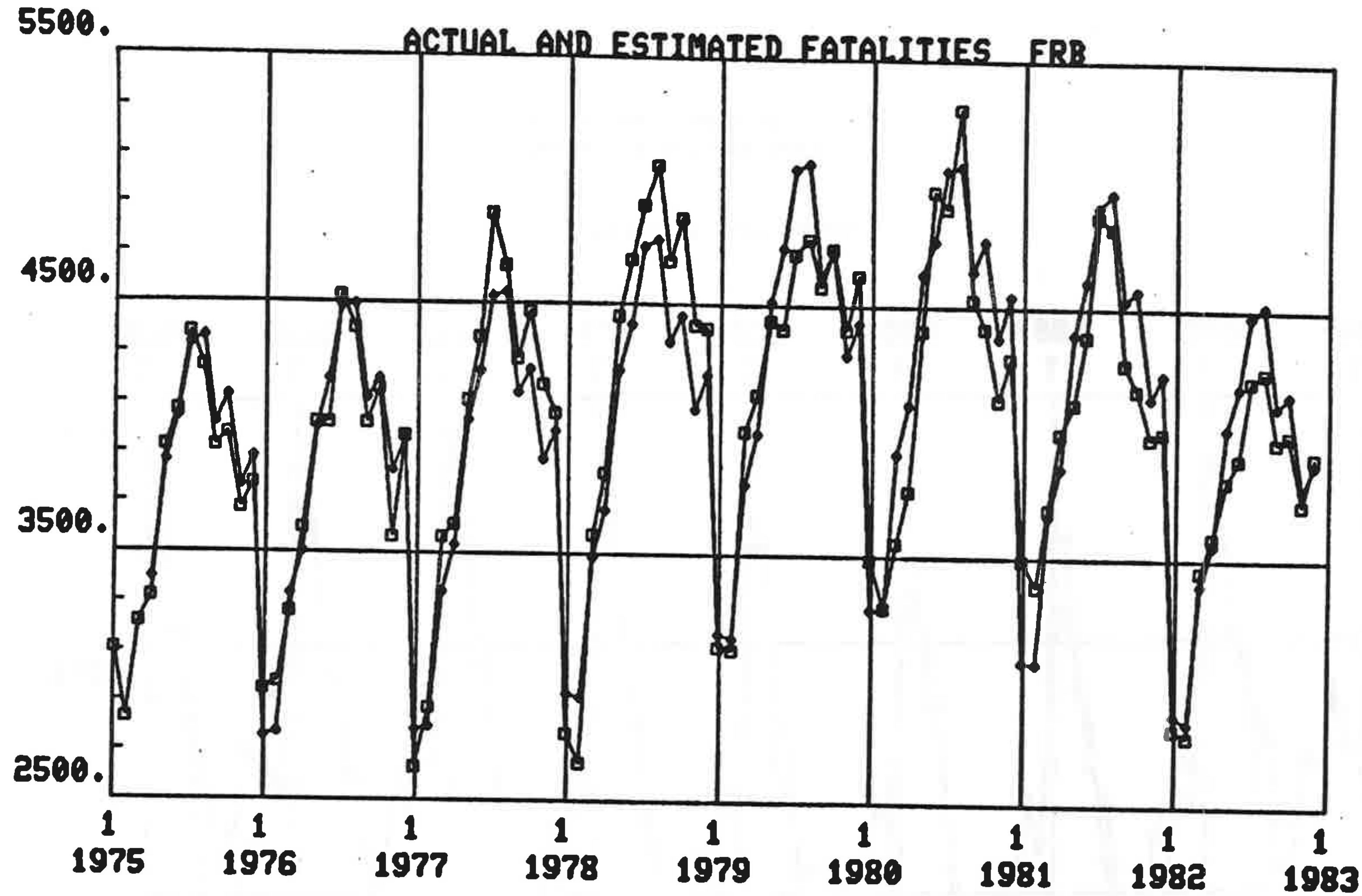


Figure 6.3: FRB Model Performance

Estimated (closed box)
Actual (opened box)

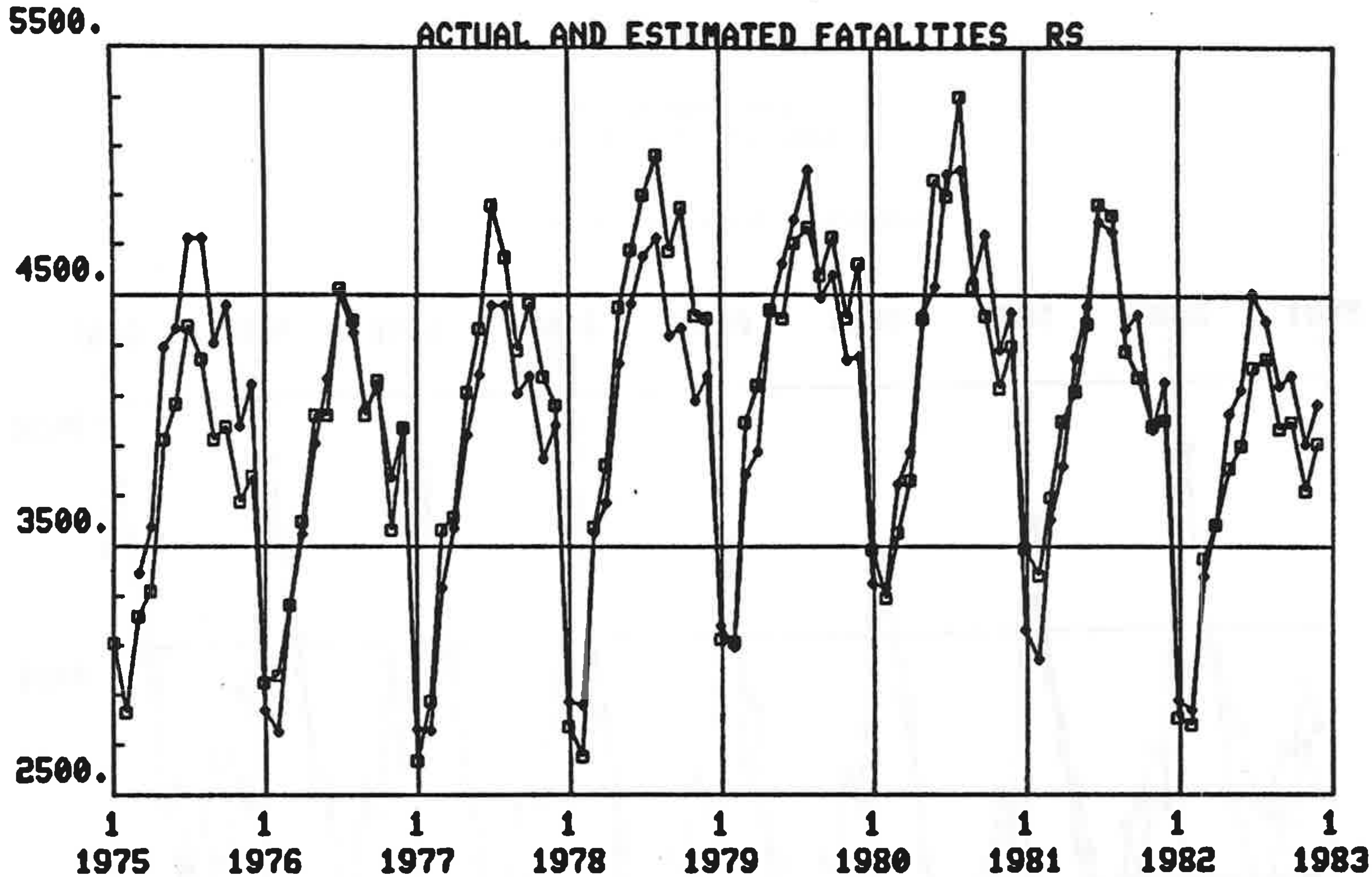


Figure 6.4: RS Model Performance

Estimated (closed box)
Actual (opened box)

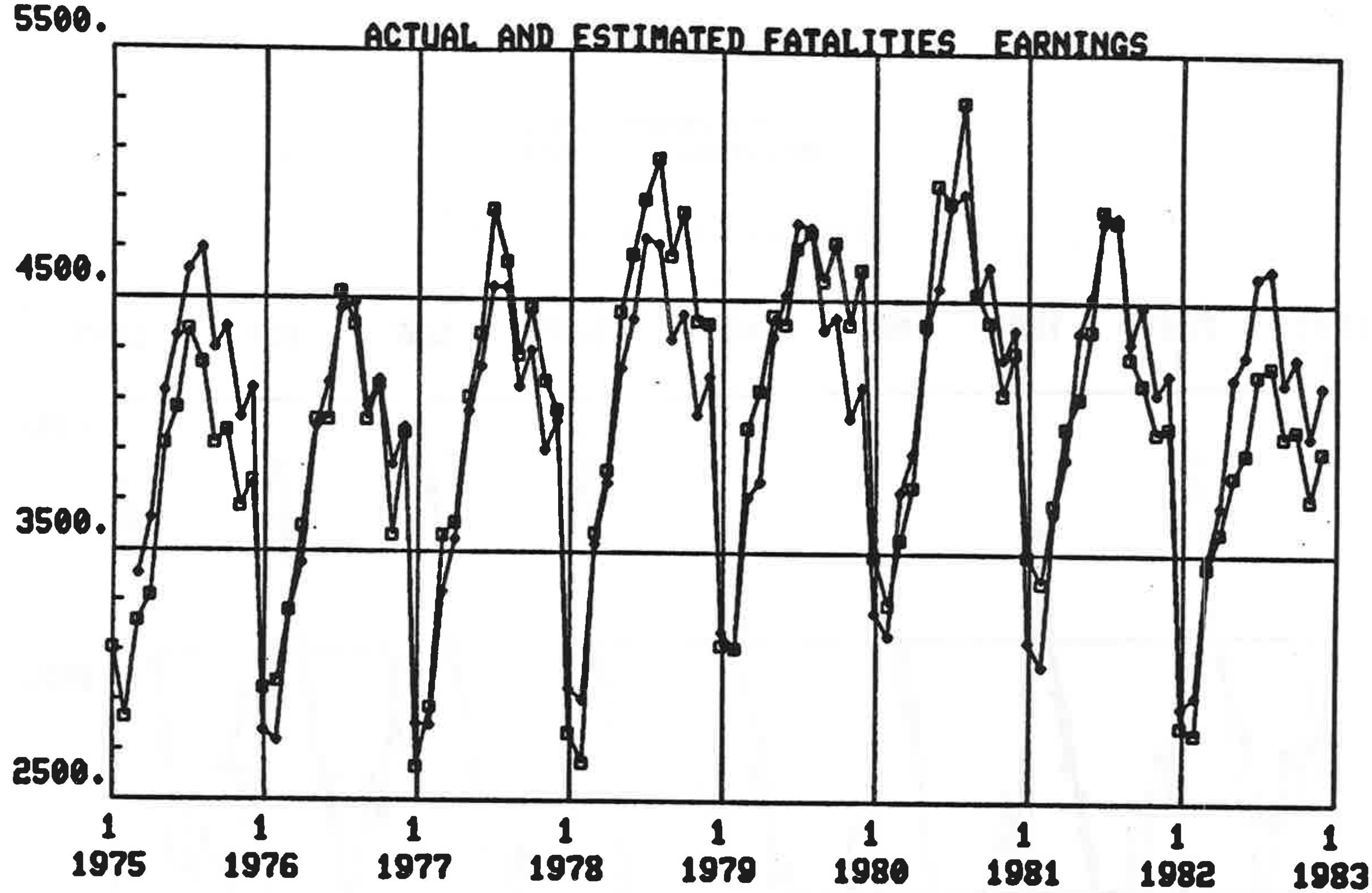


Figure 6.5: EARN Model Performance

Estimated (closed box)
Actual (opened box)

This conclusion does not exclude the possibility that there is a systematic error in the 1982 fatality estimates which is caused by an influence on fatalities not covered by any of the models. Many factors influence fatalities which we have not considered because they are not specifically "socio-economic." Safer cars, and roads, changes in restraint use, drunk driving campaigns and more available emergency medical services are some obvious examples. Further, it seems very likely that there are important economic influences on highway fatalities which are not in these models even though a preliminary search for some of these influences, discussed in Section 5.3, resulted in rejection of the five influences tested. Further work in this area is warranted.

6.2 Analysis of the VMT Model

Given that the relationship between VMT and highway fatalities is about the same in 1982 as it was back until 1975, how can a one percent increase in VMT lead to a 10 percent decrease in fatalities? In answering this question, the behavior of the model will be examined to arrive at a better understanding of the influence of VMT on fatalities. The VMT model estimated 1982 fatalities to be 6.3 percent lower than actual 1981 fatalities.

The model of fatalities (F_t) is:

$\Delta \log F_t = a + \text{seasonal dummies} + 1.413\, \Delta \log VMT_t$.*

This model was used to estimate the fatalities for each month by calculating the ratio of that month's fatalities to the prior month's fatalities:

$$F_t/F_{t-1} = e^{\log F_t} / e^{\log F_{t-1}} = e^{\log F_t - \log F_{t-1}} = e^{\Delta \log F_t}$$

Then this ratio is multiplied by the prior month's fatalities to produce the current month's fatality estimate. The process starts with the known number of

*Note that the RH01 term (see Table 5.1) in this model is mainly used to get the best estimates of the coefficients but it does influence the forecasts in January and to a much smaller degree in February. It is ignored here for the sake of simplicity.

fatalities in December and uses the change in the log of VMT for each successive month, the constant and the seasonal dummies to produce the fatality estimates for each month.

The model's forecast, aggregated over a year, can be decomposed into four components: (1) the December starting point; (2) the combined effect of the constant and the seasonal dummies; (3) the change in the aggregate level of VMT; and (4) the change in the monthly distribution of VMT. The effect of changes between 1981 and 1982 on each of these components is discussed below.* Much of this discussion is speculation aimed at understanding the behavior which the model summarizes. This speculation is a possible explanation of the model results but certainly many other interpretations are possible.

December 1981 is 6.7 percent lower than December 1980. Since the December 1981 fatality level affects the forecast for every month in 1982, this translates into a 6.7 percent lower estimate for fatalities in 1982. This December starting value summarizes the fatality series at the time, but it is subject to random fluctuations which affect the next year's forecast. So, a natural question is "Is the 6.7 percent decline in fatalities between December 1980 and 81 reasonable?" Since fatalities in 1981 were 3.3 percent lower than 1980, the decline should be at least 3.3 percent. Further, the first few months of 1981 had higher fatalities than 1980, so in order for the average to be a 3.3 percent decline the remaining months must be substantially lower than 3.3 percent below the 1980 levels. these considerations lead us to believe that December 1981 represent a reasonable summary of the fatality series at December which shows the substantial decline which occurred in the last half of 1981.

The combined effect of the constant and the seasonal dummies is to reduce fatalities by 3.3 percent. This effect is estimated by using 1981 monthly VMT to estimate fatalities in both 1981 and 1982 and comparing the aggregate estimated

*Note that these four components interact strongly with one another and the estimates of the size of each effect in 1982 is approximate. The sum of these components only approximately equals the total change in 1982.

fatality levels. This effect implies that if VMT and the distribution of VMT stay the same, fatalities are expected to fall by 3.3 percent. Further, because the coefficient of VMT in the model is 1.4, VMT would have to grow by roughly 2 percent (3.3/1.4) for fatalities to remain constant.

How could this happen? First, fatalities per VMT — the fatality rate — declines over time probably because of safety improvements in car design, roadway design, driver skill, medical treatment, and alcohol and seat belt programs. One other component of this result involves the risk distribution of driving. Consider two general types of driving. One type, routine driving, increases with increases in the labor force or the driving population; the other type is discretionary driving. Routine driving, such as commuting and shopping trips is much safer than discretionary driving such as vacations or late night trips. Since the labor force and driving population grow by roughly two percent per year, the safer, routine driving probably grows by this same percentage. If total VMT grows by less than two percent, then the share of routine, safe driving must increase and the fatal accident rate should fall. If total VMT stays constant, then the 2 percent increase in safer routine driving causes an actual reduction in total fatalities because this safer driving replaces riskier discretionary driving.

Put another way, the number of drivers increases by 2 percent per year, so unless VMT increases by more than 2 percent the average miles per driver will decrease. When mileage per driver decreases, higher risk discretionary driving is reduced first and the safest work and shopping trips are reduced last. The result is a reduction in fatalities per driver.

Between 1981 and 1982 total VMT actually increased by one percent. The effect of this increase, measured by scaling-up 1981 VMT by one percent, is to increase estimated fatalities by 1.4 percent. The 1.4 factor represents a combination of many influences. Single car fatal accidents would be expected to grow in proportion to VMT. Two-car fatal accidents would be expected to grow with the square of VMT because both cars must be exposed to the accident situation. Further, any change in VMT is also likely to affect the share of routine and discretionary driving.

TABLE 6.3

MONTHLY DISTRIBUTION OF DRIVING
(Percent of Annual Driving)

MONTH	1981	1982	82-81
January	7.57	7.14	-.43
February	7.07	7.05	-.02
March	8.28	8.20	-.08
April	8.29	8.22	-.07
May	8.66	8.77	+.11
June	8.76	8.74	-.02
July	9.04	9.14	+.10
August	9.18	9.26	+.08
September	8.42	8.48	+.06
October	8.64	8.73	+.09
November	8.04	8.13	+.09
December	8.05	8.14	+.09

The effect of the changing the monthly distribution of driving was measured by comparing the aggregate fatalities for 1982 estimated using the 1981 monthly distribution of driving scaled-up by one percent with the 1982 estimate obtained from the actual 1982 distribution of driving. The 1982 distribution of driving caused fatalities to increase by 1.5 percent. Increases in summer driving are more likely to be increases in discretionary, high risk driving while increases in winter driving are more likely to be routine driving. Table 6.3 compares the percentage of annual driving in each month for 1981 and 1982. In 1982, a larger share of driving occurred in summer months where the risk is higher.

6.3 Conclusions

The relationship between VMT and highway fatalities appears to be the same as the relationship which existed between 1976 and 1981. However, there are probably other important influences on the risk distribution of total driving which affect total fatality levels. Preliminary efforts to identify these other influences have failed, but further efforts are warranted.

An examination of the VMT model of highway fatalities in 1981 and 1982 revealed that:

- for yearly fatality levels to increase, total driving — exposure to accidents — must increase by more than 2 percent to offset both the historical decline in the fatality rate and the annual addition of safe work and shopping trips.

- when total driving does increase at more than 2 percent per year it causes fatalities to increase by 1.4 percent for each 1 percent increase in driving above the 2 percent threshold. This is because the additional driving is concentrated in riskier travel situations.

- the proportion of the travel in each month has an important influence on fatality levels because summer and fall driving is riskier.

7.0 Conclusions

The objective of this study has been to explore the influence of socio-economic factors on highway fatalities and to examine these influences in 1982 when fatalities declined by 10 percent. An initial screening of 34 socio-economic variables revealed that changes in VMT were as strongly associated with changes in highway fatalities as any other variable examined. Reservations about VMT measurement accuracy had prompted the inclusion of several variables which are closely related to VMT (gasoline sales and gasoline sales divided by miles per gallon), but neither of these variables performed better than VMT.

As a result of this screening and more demanding tests of "statistical causality" and month-to-month forecasting performance, a hypothesis about the mechanism relating socio-economic factors to highway fatalities was developed: Socio-economic factors influence fatalities either because they influence the total amount of driving or because they influence the proportion of driving by high risk groups in high risk vehicles in high risk situations. Socio-economic factors influence the amount or risk distribution of driving.

Empirical testing of this hypothesis was performed on each part (amount or distribution) independently. Personal income, retail sales, average weekly earnings per production worker, and the FRB production index were hypothesized to influence fatalities because of their influence on VMT. Only personal income and retail sales were significantly associated with fatalities at the 90% level or higher. Using "path analysis" to empirically test the path influence revealed that personal income affects fatalities solely through VMT, while retail sales affects fatalities partly through VMT and partly directly, by somehow influencing the risk distribution of driving.

Attempts to identify other factors hypothesized to affect specific types of high-risk driving were unsuccessful. A reason for this failure may be that over the eight years modeled many factors have influenced the risk distribution of driving, each factor playing an important role at different times so that no single factor explains enough of the variation which is left after the effect of the amount of total driving has been removed. Probably a more successful approach is to explore the influence of socio-economic factors on specific high-risk fatality strata, like fatalities involving teenage drivers, late night fatalities, or rural fatalities.

Models of highway fatalities based on VMT or its proxies, personal income and retail sales, predicted fatality declines of between 6 and 8.5 percent in 1982. This is reasonably close to the actual 10 percent drop and these models perform no worse in 1982 than during the 1975-1981 period. Still, there is a significant unexplained component of the variation in highway fatalities. It may represent the influence of "non-socio-economic" factors, like emergency medical service availability, automobile safety improvements, and drunk driver and restraint use campaigns, and it may represent the effects of socio-economic factors on the risk distribution of driving. Further research which identifies the effects of the economy on the risk distribution of driving would help extend the research started here and provide a more complete understanding of the connections between the economy and highway safety.

This research has established the connection between the economy, measured by personal income and retail sales, and VMT and subsequently highway fatalities. The relationship between highway fatalities and VMT is not a simple one, however. A one percent change in VMT does not cause a one percent change in highway fatalities. A better picture of the relationship is presented in Figure 7-1. The intercept on the percent change in highway fatalities axis is about -3%. Thus, no change in VMT results in a 3% decline in fatalities. This reflects safer cars and roadways and the safer driving caused by the growth in routine driving which accompanies the increases in the number of licensed drivers. The slope of the line is 1.4 so one percent change in VMT causes a roughly 1.4 percent change in fatalites. This slope reflects the higher risk nature of the marginal driving. Another aspect of this picture not in Figure 7-1, is that summer and fall driving tends to be riskier than winter and spring driving, so that when the changes in VMT occur influences the effect on fatalities.

What is 1983 like and how well does it conform to the 1982 results? Table 7-1 presents the actual highway fatalities (adjusted for working day/trading day variation) for the first six months of 1983 and the percentage errors in the forecasts using VMT, personal income and retail sales. Fatalities declined by roughly five percent over 1982 while the three models predicted a slight increase in fatalities. Further, the RMS error for these three models appears to be somewhat higher than would be expected, (based on 1975-1982 results in Table 6-2) suggesting that there may be a change in the relationship between fatalities and VMT and its surrogates. Further research on the economic influences on high-risk fatality strata may identify causes for this apparent change.

Figure 7-1

Approximate Relationship Between Changes in VMT and Changes in Highway Fatalities

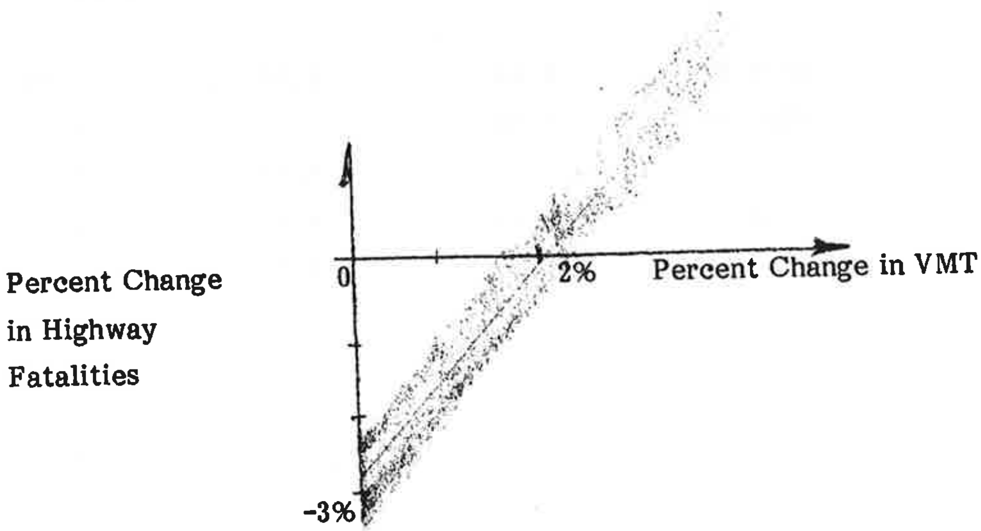


Table 7-1

Forecast Results for 1983

Month	Actual*	VMT	PI	RS
		Percent Error (Estimated - Actual)/Actual		
January	2743.53	6.48	1.52	3.00
February	2765.00	1.59	-1.55	-0.52
March	3170.72	5.67	2.85	6.56
April	3153.81	10.45	8.40	12.44
May	3735.20	6.34	5.92	7.75
June	3779.48	9.50	8.70	14.00
Total	19,348	20,675	20,252	20,835
Percent Error	---	6.82	4.67	7.29
RMS Percent Error	---	7.26	5.69	8.79
Estimated Percentage Change From 1982	---	1.56	-0.51	2.35

*Adjusted for working day/trading day variation.

APPENDIX A

VARIABLE	UNITS	UPDATE FREQUENCY	SOURCE	DEFINITION	NOTES
DRIVING ACTIVITY					
VMT	billions of miles	monthly	"Traffic Volume Trends" Federal Highway Administration	Vehicle miles traveled	1982 values are preliminary estimates
GSALES	thousands of gallons	monthly	"Monthly Motor Gasoline Reported by States" FHA	Wholesale gross gallons of gasoline reported to state tax agencies	
GMILES	thousands of miles			GSALES divided by miles per gallon	Miles per gallon values are interpolated to monthly values
STATSLS	millions (72$)	monthly	"Monthly Retail Trade" U.S. Dept. of Commerce, Bureau of the Census	Estimated monthly retail sales for gasoline service stations	Adjusted by Consumer Price Index
MPG	billions of miles per gallon			VMT/GAS	A derived miles per gallon figure. Not the same as the miles per gallon which is used in GMILES which is published annually in "Highway Statistics"
DRIVING COST					
GAS	cents per gallon (72$)	weekly	"Oil and Gas Journal"	Average weekly gasoline price per month	Average U.S. pump price for regular major-brand gasoline including taxes Used to measure variable costs.

APPENDIX A (cont'd.)

VARIABLE	UNITS	UPDATE FREQUENCY	SOURCE	DEFINITION	NOTES
CARCOST	index	monthly	Bureau of Labor Statistics	Consumer Price Index for all Urban Consumers: U.S. City Average Private Transportation	Composed of costs of new and used autombiles, gasoline and oil, tires, repairs and mantenance, insurance, and registration fees. Used to measure total costs.
INCOME AND EMPLOYMENT					
UNEMP	thousands	monthly	"Employment and Earnings," Bureau of Labor Statistics	Absolute number of unemployed	Labor force survey
UNEMP%	percent	monthly	"Employment and Earnings," Bureau of Labor Statistics	UNEMP/LF	Labor force survey
YUNEMP	thousands	monthly	"Employment and Earnings," Bureau of Labor Statistics	Absolute number of youth unemployed, age 16-19, both sexes	Labor force survey
DUNEMP	weeks	monthly	"Business Conditions Digest," U.S. Department of Commerce	Average (mean) duration of unemployment in weeks	A measure of severity of unemployment
EMP	thousands	monthly	"Employment and Earnings," Bureau of Labor Statistics	Absolute number of employed	
EMP%	percent	monthly	"Business Conditions Digest," U.S. Department of Commerce	Ration of civilian employment to total population of working age	
DI	billions(72$)	monthly	"Survey of Current Business," U.S. Department of Commerce	Disposable Income	Seasonally adjusted at annual rates

APPENDIX A (cont'd.)

VARIABLE	UNITS	UPDATE FREQUENCY	SOURCE	DEFINITION	NOTES
PI	billions (72$)	monthly	Citybank Economic Database	Personal Income	Seasonally adjusted
DIPC	billions(72$)	monthly	derived	Disposable income per capita DI/POP	
EARN	(72$)	monthly	"Monthly Labor Review," U.S. Department of Labor	Average gross earnings per production or non supervisory worker on non-agricultural payrolls expressed in earnings per week average	Adjusted by CPI
RS	millions(72$)	monthly	"Monthly Retail Trade," U.S. Department of Commerce	Total retail trade	Adjusted by CPI
FRB	index 1967 = 100	monthly	"Survey of Current Business," U.S Department of Commerce	Federal Reserve Board Index of Quantity Output (Production)	
C.IND	index	monthly	"Business Conditions Digest," U.S. Department of Commerce	Composite Index of 4 roughly coincidental indicators	Seasonally adjusted
L.IND	index	monthly	"Business Conditions Digest," U.S. Department of Commerce	Composite INdex of 12 leading indicators	Seasonally adjusted
DEMOGRAPHIC (Population and Motor Vehicle Fleet)					
POP	thousands	monthly	"Current Population Reports," U.S. Department of Commerce, Bureau of the Census	Total noninstitutional population	
LF	thousands	monthly	"Employment and Earnings," Bureau of Labor Statistics	Absolute number in labor force	

APPENDIX A (cont'd.)

VARIABLE	UNITS	UPDATE FREQUENCY	SOURCE	DEFINITION	NOTES
RD	millions	yearly	(internal document) National Highway Traffic Safety Administration	Registered drivers	monthly values inter-preted from annual values
CARREG		yearly	"Highway Statistics," Federal Highway Administration	Total number of cars registered	monthly values inter-preted from annual values
NCR		monthly	"1983 Ward's Automotive Yearbook," (R.L. Polk S Co.)	U.S. new car registration	
NTR%	percent	monthly	derived	$\frac{NTR}{(NTR+NCR)}$	
FLEETR		monthly	"Automotive Fleet"	Number of fleet registrations for all fleets of 10 or more vehicles	
FCR		monthly	"1983 Ward's Automotive Yearbook," (R.L. Polk S Co.)	Foreign car registrations	
<u>OTHER</u>					
CPI	index 1967 = 100	monthly	"CPI Detailed Report," Bureau of Labor Statistics	"Consumer Price Index for All Workers"	
HUDI	percent	monthly	"Federal Reserve Bulletin Federal Reserve	HUD mortgage interest rate in primary market	

APPENDIX A (cont'd.)

VARIABLE	UNITS	UPDATE FREQUENCY	SOURCE	DEFINITION	NOTES
HMI	percent	monthly	"Survey of Current Business," U.S. Department of Commerce	Home mortgage rates: conventional first mortgages on new home purchase (U.S. average)	
LS	millions (72$)	monthly	"Monthly Retail Trade," U.S. Department of Commerce	Sum of sales for 'liquor stores' and 'drinking places'	

APPENDIX B

RESULTS OF THE MODIFIED YEATS PROCEDURE: LAGS 0-12 MONTHS

LENGTH OF LAG IN MONTHS

VARIABLE	0	1	2	3	4	5	6	7	8	9	10	11	12
<u>Driving Activity</u>													
VMT													
Coef.	1.413	1.062	-.185	-.421	-.547	.302	.251	-.102	-.089	-.486	-.080	-.144	.626
t-Stat.	4.96	3.62	-.56	-1.32	-1.68	.93	.77	-.31	-.27	-1.48	-.24	-.42	1.86
GSALES													
Coef.	.545	.655	-.290	-.077	.126	-.132	-.127	.041	-.112	.279	-.152	.059	.415
t-Stat.	2.60	3.20	-1.38	-.36	.58	-.59	-.57	.18	-.49	1.24	-.66	.25	1.69
GMILES													
Coef.	.550	.655	-.282	-.075	.124	-.122	-.112	.048	-.098	.284	-.143	.066	.421
t-Stat.	2.63	3.23	-1.35	-.35	.57	-.55	-.50	.22	-.44	1.26	-.62	.28	.172
STATSLS													
Coef.	.273	-.178	-.098	.229	.040	-.100	.324	-.005	.007	-.150	-.246	.166	0.588
t-Stat.	1.03	-.68	-.38	.87	.15	-.37	1.21	-.02	.03	-.53	-.85	.56	2.05
MPG													
Coef.	.546	.490	.081	-.184	-.346	-.074	-.025	.079	.173	-.138	-.028	-.110	.052
t-Stat	3.19	3.02	.46	-1.04	-1.99	-.41	-.138	.43	.93	-.71	-.12	-.54	.26
<u>Driving Cost</u>													
GAS $													
Coef.	-.203	-.399	-.205	.108	.326	.291	.180	-.195	-.351	-.051	-.006	.113	.294
t-Stat.	-.83	-1.51	-.876	.454	1.40	1.24	.75	-.80	-1.44	-.19	-.02	.40	1.07
CARCOST													
Coef.	-.763	-.929	-.064	.250	.792	.626	-.067	-.837	-1.154	-.325	-.170	.100	.414
t-Stat.	-1.20	-1.50	-.10	.40	1.26	.98	-.10	-1.29	-1.77	-.46	-.22	.12	.52
<u>Income and Employment</u>													
UNEMP													
Coef.	-.055	.244	-.174	.016	.009	-.221	-.105	-.085	.208	.189	0.0	-.140	-.052
t-Stat.	-.36	1.52	-1.17	.11	.06	-1.46	-.68	-.54	1.32	1.17	0.0	-.85	-.32

RESULTS OF THE MODIFIED YEATS PROCEDURE: LAGS 0-12 MONTHS (CONT'D)

LENGTH OF LAG IN MONTHS

VARIABLE	0	1	2	3	4	5	6	7	8	9	10	11	12
UNEMP %													
Coef.	-.086	.196	-.148	.079	-.010	-.207	-.115	-.094	.224	.205	.117	-.184	-.048
t-Stat.	-.561	1.30	-.98	.48	-.061	-.134	-.74	-.58	1.40	1.25	.07	-1:09	-.29
YUNEMP													
Coef.	-.058	.149	-.070	-.024	-.068	-.178	-.126	-.003	.122	.241	.067	-.140	.012
t-Stat.	-.38	1.00	-.47	.16	-.45	-1.17	-.82	-.02	.79	1.56	.42	-.87	.07
DUNEMP													
Coef.	-.102	.112	.042	-.164	-.132	.015	.139	-.145	.129	-.014	.060	.008	-.188
t-Stat.	-.76	.85	.32	-1.23	-.99	.11	1.02	-1.05	.91	-.10	.42	.05	-1.32
EMP													
Coef.	2.109	1.106	-1.139	-1.591	1.642	-.502	.714	.863	-1.978	-1.661	-.837	3.675	.612
t-Stat.	1.75	.90	-.929	-1.26	1.30	-.39	.56	.66	-1.52	-1.26	-.62	2.80	.46
EMP%													
Coef.	2.770	-.345	-1.056	-049	2.374	1.263	.722	.409	-2.238	-2.018	-.376	4.516	1.336
t-Stat.	1.73	-.22	-.66	-.03	1.46	.77	.43	.24	-1.33	-1.19	-.21	2.68	.77
DI													
Coef.	.840	.374	-.954	-.280	-.065	.085	-.016	.524	-.223	-.467	-.230	.556	1.62
t-Stat.	1.01	.45	-1.16	-.34	.08	.10	-.02	.61	-.26	-.54	-.26	.63	1.87
DIPC													
Coef.	.644	-.035	-1.085	.102	-.295	.384	.254	.335	-.322	-.900	-.047	.901	1.246
t-Stat.	.89	-.05	-1.53	.14	-.41	.53	.35	.45	-.43	-1.21	-.06	1.20	1.69
PI													
Coef.	2.296	.709	-2.222	-.036	-.025	.913	.301	.732	.467	-2.230	-.614	.907	.528
t-Stat.	2.22	.67	-2.16	-.03	-.023	.84	.27	.65	.406	-1.97	-.53	.77	.44

RESULTS OF THE MODIFIED YEATS PROCEDURE: LAGS 0-12 MONTHS (CONT'D)

LENGTH OF LAG IN MONTHS

VARIABLE	0	1	2	3	4	5	6	7	8	9	10	11	12
EARN													
Coef	1.852	1.296	-.486	-.156	-.268	.728	-.135	-.169	.120	-.903	-.254	1.647	0.421
t-Stat	2.39	1.63	-.60	-.19	-.33	.89	-.16	-.20	.14	-1.06	-.29	1.80	.45
RS													
Coef.	.861	.186	-.585	.218	-.008	-.063	.307	-.110	.038	-.019	-.447	.144	.441
t-Stat.	4.43	.90	-2.95	1.05	-.04	-.30	1.44	-.50	.17	-.09	-2.05	.64	2.01
FRB													
Coef.	.420	.528	-.011	-.366	.503	.347	-.037	-.213	-.488	-.429	.536	.107	.328
t-Stat.	1.12	1.43	.03	-.96	1.312	.90	-.09	-.54	-1.24	-1.07	.13	.26	.80
C.IND													
Coef.	1.400	.182	-.675	.105	.484	.591	.347	-.067	-.565	-.855	-.150	.566	.803
t-Stat.	2.68	.34	-1.27	.19	.88	1.06	.67	-.12	-.98	-1.48	-.25	.95	1.36
L.IND													
Coef.	.172	-.216	-.413	.254	.500	.395	.405	.300	-.363	-.568	-.304	.481	.521
t-Stat.	.45	-.58	-1.12	.68	1.36	1.07	1.08	.78	-.95	-1.50	-.78	1.23	1.37
Demographic													
POP													
Coef.	-.128	3.23	3.685	-3.479	2.780	-3.214	-2.672	.516	1.668	5.475	-1.219	-4.822	-.572
t-Stat.	-.06	1.49	1.64	-1.52	1.22	-1.42	-1.171	.22	.72	2.40	-.51	-2.07	-.24
LF													
Coef.	3.099	3.309	-2.909	-2.143	1.820	-2.141	.641	.525	-.988	-.936	-1.117	3.718	-.564
t-Stat.	2.14	2.31	-2.03	-1.45	1.23	-1.44	.42	.34	-.62	-.56	-.69	2.36	-.35
DRIVERS													
Coef.	.101	1.618	2.491	3.384	2.526	3.490	3.940	4.595	4.589	4.732	4.951	4.463	2.438
t-Stat.	.02	.31	.47	.63	.47	.64	.71	.82	.81	.82	.84	.75	.41

RESULTS OF THE MODIFIED YEATS PROCEDURE: LAGS 0-12 MONTHS (CONT'D)

LENGTH OF LAG IN MONTHS

VARIABLE	0	1	2	3	4	5	6	7	8	9	10	11	12
CARREG													
Coef.	5.162	7.007	5.795	3.196	2.532	4.892	6.914	5.630	7.120	5.643	4.230	4.257	5.152
t-Stat.	.75	1.05	.86	46	.36	.70	.97	.78	.97	.75	.55	.55	.67
NCR													
Coef.	.055	.008	-.111	.067	.139	.017	-.091	-.056	-.022	.003	.032	.012	.031
t-Stat.	.07	.12	-1.54	.92	1.94	.22	-1.22	-.74	-.29	.04	.40	.15.	39
NTR													
Coef.	.031	.075	-.014	.042	.087	-.148	-.075	.062	.030	.054	.082	-.070	-.052
t-Stat.	.42	1.04	-.19	.55	1.16	-1.98	-.95	.78	.38	.67	1.01	.84	.63
NRT%													
Coef.	.057	.134.	195	-.063	-.120	-.305	.047	.220	.095	.086	.087	-.141	-.152
t-Stat.	.48	1.17	1.71	-.54	-1.03	-2.69	.391	.88	.79	.71	.70	-1.13	-1.23
FLEETR													
Coef.	.002	-.025	.024	.002	-.016	.004	.-33	.035	-.-26	-.061	.070	-.050	.006
t-Stat.	.04	-.64	.60	.04	-.39	.104	.80	.85	-.63	-1.49	1.73	-1.19	.14
FCR													
Coef.	-.057	-.010	.60	-.104	-.004	-.033	.005	.094	.017	.025	-.018	-.068	-.029
t-Stat.	-1.00	-.18	1.08	-1.83	-.07	-.56	.09	1.55	.279	.40	-.28	-1.06	-.45
Other CPI													
Coef.	-.071	-.617	.836	.316	.192	-.099	-1.150	-.233	-1.226	-1.031	-.610	-1.812	-1.500
t-Stat.	-.05	-.46	.61	.23	.14	-.07	-.80	-.15	-.84	-.70	-.39	-1.15	-.96
HUDI													
Coef.	-.326	-.259	.152	.166	.051	.038	-.162	-.277	-.262	.011	.295	.101	-.042
t-Stat.	-2.43	-1.93	1.10	1.17	.35	.25	-1.07	-1.87	-1.74	.07	1.90	.64	-.27

RESULTS OF THE MODIFIED YEATS PROCEDURE: LAGS 0-12 MONTHS (CONT'D)
LENGTH OF LAG IN MONTHS

VARIABLE	0	1	2	3	4	5	6	7	8	9	10	11	12
HMI													
Coef.	-.243	.212	-.080	.141	.004	-.107	-.594	-.40	.233	.291	.005	-.235	-.160
t-Stat	-1.01	.88	-.32	.53	.02	-.39	-2.22	-1.45	.82	1.02	.02	-.80	-.54
LS													
Coef.	.495	.077	-.208	.257	-.238	-.014	.188	-.265	.261	.024	-.405	.254	.127
t-Stat.	2.63	.39	-1.06	1.31	-1.24	-.067	.92	-1.30	1.25	.11	-1.94	1.19	.590

APPENDIX C: Estimated Regression Equations for VMT and its Surrogates

The estimated coeffficients, t-statistics, and other statistics on the overall fit of the regression equations used in Section 6.0 are presented in this appendix. Table C.1 defines the terms used in the regression output which follows.

C.1 DEFINITIONS OF REGRESSION OUTPUT

DEL	first difference operator
LOG	natural logarithm
NOB	number of observations used in regression
NOVAR	number of coefficients or, in the equations presented here, the number of variables minus one
RANGE	dates for which NOB pertains
RSQ	R-squared
CRSQ	corrected R-squared
F	F-test for R-squared
SER	standard error of the regression
SSR	sum of the squared residuals
DW	Durbin-Watson statistic
COND(X)	condition matrix, a measure of multicollinesity between independent variables
RH01	autocorrelation correction factor for first-order correction
COEF	coeffieient
VALUE	value of coefficient
STER	standard error of coefficient estimate
T-STAT	t-statistic

VARIABLES

TFF.A	Total highway fatalities adjusted for working day/trading day variation
SM1,...,SM11	Dummy variables to measure difference in fatalities from average month. SM1 equals deviation of January from average month's value. The number part of the variable name, 1, 2, 3, ..., 11, represents the month. December value is derived from other eleven month's values
VMT.M	Monthly vehicle miles traveled
EARNR.M	Real earnings per production or nonsupervisory worker
RS.M	Monthly total retail sales in constant dollars
FRB.M	Monthly Federal Reserve Board Index of Industrial Production
PI	Real total personal income

Figure C.1: VMT Prediction Equation

```
   DEL(1 : LOG(TFF.A)) = B0+B1*DEL(1 : SM1)+B2*DEL(1 : SM2)+B3*DEL(1 : SM3)+B4*DEL(1 : SM4)+B5*DEL(1 : SM5)+B6*DEL(1 : SM6)+B7*DEL(
1 : SM7)+B8*DEL(1 : SM8)+B9*DEL(1 : SM9)+B10*DEL(1 : SM10)+B11*DEL(1 : SM11)+B13*DEL(1 : LOG(VMT.M))

NOB = 71      NOVAR = 13
RANGE =  1975 2 TO 1980 12
RSQ =   0.91082      CRSQ =   0.89237      F(12/58) =    49.366
SER =   0.0435       SSR =     0.110       DW(0) =  2.08      COND(X) =    25.87

GLS PARAMETERS

RHO1   -0.2612
```

COEF	VALUE	ST ER	T-STAT
B0	-0.00276	0.00416	-0.66324
B1	-0.10574	0.04300	-2.45901
B2	-0.06588	0.05523	-1.19284
B3	-0.10985	0.01614	-6.80480
B4	-0.06050	0.01602	-3.77731
B5	-0.01490	0.02347	-0.63472
B6	0.01589	0.02550	0.62314
B7	0.01927	0.03783	0.50947
B8	0.00469	0.04180	0.11218
B9	0.07170	0.01615	4.43841
B10	0.06002	0.01926	3.11734
B11	0.07784	0.02101	3.70557
B13	1.42564	0.32652	4.36619

Figure C.2: PI Prediction Equation

```
    DEL(1 : LOG(TFF.A)) = B0+B1*DEL(1 : SM1)+B2*DEL(1 : SM2)+B3*DEL(1 : SM3)+B4*DEL(1 : SM4)+B5*DEL(1 : SM5)+B6*DEL(1 : SM6)+B7*DEL(
1 : SM7)+B8*DEL(1 : SM8)+B9*DEL(1 : SM9)+B10*DEL(1 : SM10)+B11*DEL(1 : SM11)+B13*DEL(1 : LOG(PI))

NOB = 71      NOVAR = 13
RANGE =  1975 2 TO 1980 12
RSQ =    0.88861      CRSQ =    0.86556      F(12/58) =    38.556
SER =    0.0487       SSR =      0.138       DW(0) =  1.97     COND(X) =     7.11
```

GLS PARAMETERS

RHO1 -0.2712

COEF	VALUE	ST ER	T-STAT
B0	-0.00667	0.00585	-1.13999
B1	-0.28116	0.01812	-15.51450
B2	-0.29595	0.01792	-16.51770
B3	-0.12076	0.01782	-6.77789
B4	-0.06491	0.01792	-3.62118
B5	0.06867	0.01817	3.77841
B6	0.10919	0.01793	6.09056
B7	0.16729	0.01768	9.46060
B8	0.17216	0.01769	9.72984
B9	0.08400	0.01774	4.73612
B10	0.10453	0.01787	5.84963
B11	0.01411	0.01807	0.78063
B13	2.31106	1.23930	1.86480

Figure C.3: FRB Prediction Equation

```
1:  DEL(1 : LOG(TFF.A)) = B0+B1*DEL(1 : SM1)+B2*DEL(1 : SM2)+B3*DEL(1 : SM3)+B4*DEL(1 : SM4)+B5*DEL(1 : SM5)+B6*DEL(1 : SM6)+B7*DEL(
1 : SM7)+B8*DEL(1 : SM8)+B9*DEL(1 : SM9)+B10*DEL(1 : SM10)+B11*DEL(1 : SM11)+B13*DEL(1 : LOG(FRB.M(-1)))

NOB = 70      NOVAR = 13
RANGE =  1975 3 TO 1980 12
RSQ =    0.88798      CRSQ =    0.8644       F(12/57) =    37.653
SER =    0.0485       SSR =      0.134       DW(0) =  2.05      COND(X) =      9.65

GLS PARAMETERS

RHO1   -0.2131
```

COEF	VALUE	ST ER	T-STAT
B0	-4.54021E-04	0.00517	-0.08783
B1	-0.27921	0.02199	-12.69810
B2	-0.28104	0.02107	-13.33760
B3	-0.11849	0.01864	-6.35727
B4	-0.06996	0.01878	-3.72514
B5	0.06343	0.01803	3.51817
B6	0.10682	0.01803	5.92365
B7	0.16187	0.02014	8.03681
B8	0.18911	0.02510	7.53497
B9	0.08441	0.01807	4.67227
B10	0.09341	0.02196	4.25367
B11	0.00533	0.02104	0.25325
B13	0.43183	0.47878	0.90192

Figure C.4: RS Prediction Equation

```
    DEL(1 : LOG(TFF.A)) = B0+B1*DEL(1 : SM1)+B2*DEL(1 : SM2)+B3*DEL(1 : SM3)+B4*DEL(1 : SM4)+B5*DEL(1 : SM5)+B6*DEL(1 : SM6)+B7*DEL(
1 : SM7)+B8*DEL(1 : SM8)+B9*DEL(1 : SM9)+B10*DEL(1 : SM10)+B11*DEL(1 : SM11)+B13*DEL(1 : LOG(RS.M))

NOB = 71      NOVAR = 13
RANGE =  1975 2 TO 1980 12
RSQ =    0.90317     CRSQ =    0.88314      F(12/58) =     45.084
SER =    0.0458      SSR =      0.121       DW(0) =  2.01      COND(X) =     16.14

GLS PARAMETERS

RHO1   -0.2985
```

COEF	VALUE	ST ER	T-STAT
B0	-6.11356E-04	0.00421	-0.14525
B1	-0.18851	0.03116	-6.04975
B2	-0.18739	0.03557	-5.26871
B3	-0.11323	0.01674	-6.76599
B4	-0.06075	0.01672	-3.63337
B5	0.03904	0.01762	2.21500
B6	0.08912	0.01694	5.25955
B7	0.16370	0.01652	9.90708
B8	0.15459	0.01735	8.90839
B9	0.10806	0.01784	6.05834
B10	0.09406	0.01701	5.52855
B11	0.00125	0.01742	0.07200
B13	0.76972	0.22163	3.47302

Figure C.5: EARN Prediction Equation

```
    DEL(1 : LOG(TFF.A)) = B0+B1*DEL(1 : SM1)+B2*DEL(1 : SM2)+B3*DEL(1 : SM3)+B4*DEL(1 : SM4)+B5*DEL(1 : SM5)+B6*DEL(1 : SM6)+B7*DEL(
1 : SM7)+B8*DEL(1 : SM8)+B9*DEL(1 : SM9)+B10*DEL(1 : SM10)+B11*DEL(1 : SM11)+B13*DEL(1 : LOG(EARNR.M))

NOB = 71      NOVAR = 13
RANGE =  1975 2 TO 1980 12
RSQ =    0.88371      CRSQ =    0.85965      F(12/58) =    36.729
SER =    0.0490       SSR =      0.139       DW(0) =  1.99     COND(X) =     9.91

GLS PARAMETERS

RHO1   -0.2000
```

COEF	VALUE	ST ER	T-STAT
B0	0.00172	0.00498	0.34637
B1	-0.26442	0.02112	-12.52190
B2	-0.28330	0.02041	-13.88110
B3	-0.11104	0.01942	-5.71864
B4	-0.05151	0.02177	-2.36602
B5	0.07245	0.01971	3.67537
B6	0.10161	0.01821	5.57972
B7	0.16424	0.01847	8.89060
B8	0.16563	0.01893	8.75193
B9	0.07010	0.02058	3.40667
B10	0.09406	0.02023	4.65008
B11	0.01075	0.01915	0.56129
B13	1.61846	1.03960	1.55682

REFERENCES

1. Land, K. C. and McMilien, M. M., "A Macrodynamic Analysis of Changes in Mortality Indexes in the Unitied States, 1946-75: Some Preliminary Results," Social Indicators Research, 1980. D. Reidel Publishing Co.

2. Peltzman, S., "The Effects of Automobile Safety Regulation," Journal of Political Economy, Vol. 83, No. 4, 1975.

3. Stein, M. M., "Accident Trend Monitoring and Exploratory Analysis," Abt Associates Inc., (revised edition) January 4, 1982.

4. Joksch, H. C., "An Accident Trend Model," Center for the Environment and Man, Incorporated, prepared for The National Highway Traffic Safety Administration, March 1975.

5. Leveu, Alfred J., "The Sensitivity of Work and Non-Work Travel to Energy Shortages or Price Rises," Preliminary Research Report 201, New York State Department of Transportation, Albany, New York, July 1981.

6. Granger, C. W. J., "Investigating Causal Relations by Economietric Models and Cross Spectral Methods," Econometrica, Vol. 83, No. 4, 1969, pp. 424-438.

7. Hsiao, C., "Autoregressive Modeling of Canadian Money an Income Data," Journal of American Statistical Association, Vol. 74, 1979, pp. 553-560.

8. Wang, G. H., "A Study of Economic Indicators in Rail Rreight Traffic Cycles 1950-76," Transportation Research, Vol. 15, 1981, pp. 391-405.

9. Yeats, A. J., "An Alternative Approach to the Identification of Leading Indicators," Business Economics, Vol. 7, No. 4, 1972, pp. 7-11.

10. Wang, G. H., "An Intervention Analysis of Interrupted Urban Transit Time Series: Two Case Studies," Proceedings of Business and Economic Statistics Section, American Statistical Association, 1981, pp. 424-429.

11. Wright, S., "The Method of Path Coefficients," Annals of Mathematical Statistics, Vol. 5, 1934, pp. 161-215.

12. Tukey, J. W., "Causation, Regression, and Path Analysis." contained in Kempthorne, O., et al, editors, Statistics and Mathematics in Biology, Ames, Iowa State College Press, 1954, pp. 34-45.

13. Ducan, O. D., "Path Analysis: Sociological Examples," American Journal of Sociology, Vol. 72, 1966, pp. 1-16.