## DEPARTMENT OF TRANSPORT AND COMMUNICATIONS

## FEDERAL OFFICE OF ROAD SAFETY

## DOCUMENT RETRIEVAL INFORMATION

Report No.	Date	Pages	ISBN	ISSN
CR 106	July 1992		0 642 51172 1	0810 770X

## Title and Subtitle

Factors affecting fatal road crash trends.

## Author(s)

Pettitt, A.N. Haynes, M.A. Low Choy, S.

## Performing Organisation (Name and Address)

Statistical Consulting Unit School of Mathematics Queensland University of Technology

## Sponsor

Federal Office of Road Safety GPO Box 594 CANBERRA ACT 2601

## Available from

Federal Office of Road Safety GPO Box 594 CANBERRA ACT 2601

## Abstract

This report describes the influence of non-road safety factors on the level of fatal road crashes.

Economic, social and meteorological factors were analysed as independent processes, each capable of influencing the number of fatal crashes. Various statistical models were fitted to determine the power of each factor to predict crash trends.

This report's findings are summarised in report CR 109.

## Keywords

Fatal road crash trends; economic factors; social factors; meteorological factors; non-road safety factors; predicting road crash trends.

## NOTES:

- (1) FORS Research reports are disseminated in the interests of information exchange.
- (2) The views expressed are those of the author(s) and do not necessarily represent those of the Commonwealth Government.
- (3) The Federal Office of Road Safety publishes four series of research report
  (a) reports generated as a result of research done within the FORS are published in the OR series;

(b) reports of research conducted by other organisations on behalf of the FORS are published in the CR series.

(c) reports based on analyses of FORS' statistical data bases are published in the SR series.

(d) minor reports of research conducted by other organisations on behalf of FORS are published in the MR series.

# Factors affecting fatal road crash trends

This study investigates the effect of various non-road safety factors on the level of fatal road crashes. Steps were taken to develop equations capable of predicting future levels.

## Factors affecting fatal road crash trends (CR106)

This report is a single volume with two distinct parts:

1. Literature Review of Explanatory and Predictive Models for the Number of Fatal Road Crashes

> A detailed literature review of factors which have been investigated for their ability to explain or predict the number of fatal road crashes.

2. Explanatory and Predictive Models for the Number of Fatal Road Crashes.

Describes steps taken by this study to develop and test various statistical models. These models were fitted to various economic, social and meteorological factors to determine the power of each factor to predict fatal road crash trends.

**N.B.** A short summary of the main findings of this work is also available in the following separate report:

Factors affecting fatal road crash trends: Summary Report.

(CR109)

Part 1

Literature Review of Explanatory and Predictive Models for the Number of Fatal Road Crashes

## Contents

1	Inti	roduction	6
	1.1	Overall	6
	1.2	Road crash data	7
	1.3	Aggregation of data	7
	1.4	In-depth accident studies	8
	1.5	Before-and-after studies	8
	1.6	Explanatory models	8.
	1.7	Predictive models	9
		•	
Ι	Cł	noice of response variable	10
2	Ern	1051174	10
-	21	Relating road deaths to motorization	10
	2.1	Smeed's equation	10
	2.2	Motorization level	19
	2.5		12
	2.4	Pick comparation	12
	2.5		13
	2.0		14
	2.1		15
11	С	hoice of explanatory variables	16
11 3	C Bef	hoice of explanatory variables ore-and-after studies	16 16
11 3	C Bef 3.1	hoice of explanatory variables ore-and-after studies Simple before/after comparison method	16 16 16
11 3	C Bef 3.1 3.2	hoice of explanatory variables ore-and-after studies Simple before/after comparison method	16 16 16 17
11 3	<b>Bef</b> 3.1 3.2 3.3	Choice of explanatory variables      core-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures	16 16 17 17
11 3 4	Befa 3.1 3.2 3.3 Reg	bore-and-after studies Simple before/after comparison method	16 16 17 17
11 3 4	C Befa 3.1 3.2 3.3 Reg 4.1	Choice of explanatory variables      core-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      cression analysis      Peltzmann's initial model	16 16 17 17 18 18
11 3 4	C Bef 3.1 3.2 3.3 Reg 4.1 4.2	Choice of explanatory variables      core-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      cression analysis      Peltzmann's initial model      Criticisms	16 16 17 17 18 18 18
11 3 4	Bef 3.1 3.2 3.3 Reg 4.1 4.2 4.3	Choice of explanatory variables      Ore-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Cression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies	16 16 17 17 18 18 18 19 21
11 3 4	Befa 3.1 3.2 3.3 Reg 4.1 4.2 4.3 4.4	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Cression analysis      Peltzmann's initial model      Criticisms      Problems with regression models	16 16 17 17 17 18 18 19 21 22
11 3 4	C Bef 3.1 3.2 3.3 Reg 4.1 4.2 4.3 4.4	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Cression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies      Problems with regression models	16 16 16 17 17 18 18 18 19 21 22 23
11 3 4 5	C Bef 3.1 3.2 3.3 Reg 4.1 4.2 4.3 4.4 Eco	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Tression analysis      Peltzmann's initial model      Criticisms      Problems with regression models      nomic relationships      Becession impact	16      16      17      17      18      19      21      22      23      23
11 3 4 5	<b>Bef</b> 3.1 3.2 3.3 <b>Reg</b> 4.1 4.2 4.3 4.4 <b>Eco</b> 5.1 5.2	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Tression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies      Problems with regression models      nomic relationships      Recession impact	16 16 16 17 17 18 18 19 21 22 23 23 24
11 3 4 5	C Bef 3.1 3.2 3.3 Reg 4.1 4.2 4.3 4.4 Eco 5.1 5.2 5.3	Choice of explanatory variables      ore-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Arguments initial model      Criticisms      Crosssectional studies      Problems with regression models      nomic relationships      Recession impact      Indirect economic effects	16 16 17 17 18 18 19 21 22 23 23 24 24
11 3 4 5	C Bef 3.1 3.2 3.3 Reg 4.1 4.2 4.3 4.4 Eco 5.1 5.2 5.3 5.4	Choice of explanatory variables      ore-and-after studies      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Arguments against the effectiveness of countermeasures      cression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies      Problems with regression models      nomic relationships      Recession impact      Indirect economic effects      Vnemployment      Availability of medical facilities	16 16 16 17 17 18 18 19 21 22 23 23 24 24 24 26
11 3 4 5	<b>Bef</b> 3.1 3.2 3.3 <b>Reg</b> 4.1 4.2 4.3 4.4 <b>Eco</b> 5.1 5.2 5.3 5.4 5.5	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Stression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies      Problems with regression models      nomic relationships      Recession impact      Unemployment      Availability of medical facilities	16 16 16 17 17 18 18 19 21 22 23 23 24 24 24 26 27
11 3 4 5	<b>Bef</b> 3.1 3.2 3.3 <b>Reg</b> 4.1 4.2 4.3 4.4 <b>Eco</b> 5.1 5.2 5.3 5.4 5.5 5.6	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Cression analysis      Peltzmann's initial model      Criticisms      Crosssectional studies      Problems with regression models      Nomic relationships      Recession impact      Indirect economic effects      Unemployment      Availability of medical facilities      Investment into traffic safety facilities	16      16      17      17      18      19      21      22      23      23      24      24      26      27      28
11 3 4 5	<b>Bef</b> 3.1 3.2 3.3 <b>Reg</b> 4.1 4.2 4.3 4.4 <b>Eco</b> 5.1 5.2 5.3 5.4 5.5 5.6	Choice of explanatory variables      Simple before/after comparison method      More sophisticated techniques      Arguments against the effectiveness of countermeasures      Arguments initial model      Criticisms      Crosssectional studies      Problems with regression models      Indirect economic effects      Unemployment      Availability of medical facilities      Investment into traffic safety facilities	16 16 16 17 17 18 18 19 21 22 23 23 24 24 24 24 26 27 28 28

		5.6.0 New and a sisterations	•••	
	57	Discontent	29	
	5.1	Vouth	30	
	0.0	10utn	30	
6	Oth	Other explanatory variables		
	6.1	Level of vehicle inspection	31	
	6.2	Rural and urban speed limits	32	
	6.3	Variability of speeds	32	
	6.4	Public transport fares	33	
	6.5	Airline deregulation	34	
	6.6	Holiday effects	35	
	6.7	Using a number of explanatory variables	35	
	6.8	Use of seatbelts	36	
	6.9	Blood alcohol concentration (BAC) levels	36	
	6.10	Urban planning	37	
	6.11	Weather as an explanatory variable	39	
		6.11.1 Local effects studies	39	
		6.11.2 Global effects models	40	
		6.11.3 Generalised climatic variables	40	
		6.11.4 More recent applications of generalised indices	41	
	6.12	Fuel prices	41	
		6.12.1 Predicting Fuel Prices	41	
7	$\mathbf{Tim}$	e series analysis	43	
	7.1	Comparison of uncorrelated normal and ARIMA error structures	43	
		7.1.1 The regression model	43	
		7.1.2 Regression results	44	
		7.1.3 Model with ARIMA error term	45	
		7.1.4 Comparison	45	
	7.2	Intervention analysis	45	
	7.3	Identification of unknown intervention times in time series	45	
	7.4	ARIMA model, with explanatory variables	47	
	7.5	Structural time series models	47	
•	041	as statistical models	51	
0		er statistical models	51	
	0.1 0.0	Learning-incory models	21 91	
	0.2	A systems-based model	52	
	ð.3	Poisson distribution	04 52	
	ð.4	Poisson regression	03 52	
	8.5	Utner distributions	53	
	8.6	Epidemiological approach	041 54	
		8.0.1 International aggregates	04 54	
		8.6.2 Comparing international statistics	04 55	
		8.6.3 Risk and exposure	22	

-

	8.7	Discriminant analysis	55
9	List	of explanatory variables	56

.

.

# glossary

Abbrev.	Definition
ABS	Australian Bureau of Statistics
ARIMA	Autoregressive, Integrated, Moving Average
	model
BAC	Blood Alcohol Level
BAT	Blood Alcohol Testing unit
BR	British Rail
CPI	Consumer Price Index
DUI	Driving Under the Influence (of drugs or alcohol)
GDP	Gross Domestic Product
GNP	Gross National Product
$I[\boldsymbol{x}]$	If $z$ is true, then this evaluates to 1, otherwise 0.
iid	independent and identically distributed (random
	variables)
$\mathbf{LT}$	London Transport
LT KSI	London Transport Killed and Seriously Injured
LT KSI NBER	London Transport Killed and Seriously Injured National Bureau of Economics Research (US)
LT KSI NBER PCGNP	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product
LT KSI NBER PCGNP PSV	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles
LT KSI NBER PCGNP PSV pti	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants
LT KSI NBER PCGNP PSV pti R <sup>2</sup>	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression
LT KSI NBER PCGNP PSV pti R <sup>2</sup>	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression model compared to a null model
LT KSI NBER PCGNP PSV pti R <sup>2</sup> RDU	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression model compared to a null model Restraint Device Usage
LT KSI NBER PCGNP PSV pti R <sup>2</sup> RDU RPI	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression model compared to a null model Restraint Device Usage Retail Price Index
LT KSI NBER PCGNP PSV pti R <sup>2</sup> RDU RPI US	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression model compared to a null model Restraint Device Usage Retail Price Index United States
LT KSI NBER PCGNP PSV pti R <sup>2</sup> RDU RPI US VKT	London Transport Killed and Seriously Injured National Bureau of Economics Research (US) Per Capita Gross National Product Passenger-Service Vehicles Per Thousand Inhabitants amount of variation explained by a regression model compared to a null model Restraint Device Usage Retail Price Index United States Vehicle Kilometres Travelled

L

## 1 Introduction

This paper reviews both Australian and worldwide literature in the area of road and traffic safety. The investigation is focussed on road crash fatalities and fatal road crashes, and how these have been statistically modelled by researchers. Given that the variables of interest are road crash fatalities and fatal road crashes, it is then necessary to decide the following:

- 1. What is the form of the response variable? Is a ratio of fatalities to population or vehicles registered a more meaningful response variable?
- 2. Which explanatory variables should be included in the model, as suggested by theory, in such areas as road safety, socioeconomics, and psychology?
- 3. What type of statistics model and error structure is appropriate?

The three major requirements are addressed in three sections.

- §1 Choice of response variable
- §2 Choice of explanatory variable(s)
- §3 Choice of statistical model (also addressed in §1 and §2.)

#### 1.1 Overall

Overall, the literature in the area of the analysis of road fatalities seems to be concentrated on two major types of models, local effects models and global effects models. The characteristics of these two types of models are tabulated below.

Characteristic of model	Local effects models (micro-level)	Global effects models (macro-level)
applications	indepth analyses	macro-models, regres- sion, time-series analysis
focus of model	local impacts present at each particular fatality	global impacts on the to- tal number of fatalities
examples of model effects	emotional, physical, and psychological state of driver; vehicle character- istics; road design; road conditions	state of the economy, fuel prices; average weather conditions; number of ve- hicles on the road, num- ber of young drivers.
effects of coun- termeasures	effect of installation of new set of traffic lights	effect of new legislation

Hakim & Shefer review 15 of the more recent papers, which use a wide range of explanatory variables, in models which are mostly based on regression models. Some of the models they cover are discussed elsewhere in this report: Eshler (1977), Fridstrøm (1989), Hautzinger (1986), Joksch (1984), Loeb (1987), Partyka (1984), Peltzmann (1975) and Zlatoper (1984), and one paper with a time series approach, Wagenaar (1984). The authors discuss some 'issues and probelms' with analysing global effects.

#### 1.2 Road crash data

Many developing countries have only recently, or are still in the process, of determining the structure of the data on fatal road crashes to be collected. Most European countries, America and Australia have passed this milestone, and now devote much of their research effort to analysing this information. This change in collection itself could be changing what we are measuring and thus produce spurious results. Fridstrøm and Ingebrigtsen (1991) incorporated the effect of changing accident reporting procedures into their models for traffic safety in Norway. Unfortunately this effect was confounded by concurrent changes in legislation.

Analyses of road crash data are usually one of following types, in increasing order of sophistication:

#### 1. aggregation of data

- (a) numbers of road crash fatalities and fatal road crashes
- (b) road crashes according to country, state within country, road user type, age, sex, etc
- 2. in-depth accident studies
- 3. before-and-after studies of the effectiveness of countermeasures
- 4. using various explanatory variables or factors to explain the variation in the numbers of fatal road crashes or fatalities
- 5. using time-series models to predict future numbers of fatal road crashes or fatalities

#### 1.3 Aggregation of data

Similar to other countries, Australia has a Bureau of Statistics (ABS) which publishes monthly national figures on: the number of road crashes, the number of fatal road crashes, the number of fatalities resulting from road crashes. Corresponding figures for the different states are not always available; only Victoria, Queensland, Tasmania breakdowns are supplied.

ì.

In order to model the number of fatalities resulting from road crashes and the number of fatal crashes, our study relies heavily on these aggregates. Thus the degree of disaggregation available (by state) determines, to some extent, the scope of our study.

Care is required when using aggregated accident data, since different groups of crash victims might not behave similarly. For example, compare multi-vehicle and single vehicle crashes, or vehicle-occupant and pedestrian accidents. (See Hakim et al (1991), among others for more discussion.)

#### 1.4 In-depth accident studies

In-depth or local-effects accident studies try to relate the specific features of an accident to the severity of the injuries in order to locate primary causes and then to finally eliminate them. Thus fatal accidents are a natural focus for determining the major contributing factors to serious road crashes. These studies concentrate on the local factors influencing the crash, such as road geometry, signage, roadside objects, BAC level of driver, age/sex of driver, weather conditions at the scene of the accident, and the time, day, and month of accident.

Most of the research effort has gone into these types of studies, worldwide. However, due to their specific nature, these models are not relevant to our study.

#### 1.5 Before-and-after studies

Before-and-after studies are generally conducted by governmental departments to determine the impact of countermeasures either at a local or a global level. They may investigate the effectiveness of some global countermeasure such as new legislation, by state/county or entire nation; or the effectiveness of some local countermeasure, such as the installation of some new traffic safety device (traffic lights, new road design, etc).

There is a multitude of papers discussing the success or otherwise of various legislative measures. Seat-belt legislation, drink-driving laws, age of licence, and drinking age are some of the main measures considered over the last two decades.

In particular, there are a number of papers (eg Peltzmann (1975)) which have concluded that the effect of countermeasures has been swamped by a steady downward trend in fatalities during the last two decades.

#### 1.6 Explanatory models

These models have been considered in international research much more than time series models, but considerably less than before-and-after studies or indepth accident analyses. According to Hakim et al (1991), explanatory models have two main advantages:

• They provide understanding about the causes of accidents.

• They provide a basline for evaluating the effectiveness of countermeasures.

There are several types of models identified in the literature:

- 1. longitudinal models, often called time series models
- 2. crosssectional models
- 3. pooled crosssectional and longitudinal, or spacio-temporal models

Hakim et al (1991) pointed out the advantages and disadvantages of these different types. It is hard to control for geographical differences, such as climate, lifestyle, etc, in cross-sectional analyses. With annual longitudinal data, at least 30 years' worth of observations are required to provide the minimum database, and a large time trend appears to dominate most of the variability in annual series. With monthly longitudinal data, the number of years of data required is much smaller but less countries/states collect data on a monthly basis than on an annual basis.

#### 1.7 Predictive models

Often regression models have been used to predict fatality rates or numbers in future years, eg (Partyka 1984, 1991). Hakim et al (1991) gave a good review of models used for prediction, and the problems associated with using them. They considered both regression-type models assuming an independent error structure, and ARIMA or structural models which do not assume independent error structures.

Relatively little has been done in this area, even overseas. Most of the work in using time series models (with autocorrelation error structures as opposed to regressions with trend and monthly effects) was prompted by a need to model the impact of legislation, such as seatbelt or speed reduction.

# Part I Choice of response variable

## 2 Exposure

The response variable used in an analysis is often the raw number of fatalities, fatal road crashes or accidents. Two other forms of the reponse variable, which are thought to be especially useful when comparing fatality rates in different geographic regions/counties:

- 1. fatality rate per head of population: especially from a health or epidemiological point of view
- 2. fatality rate per vehicle mile travelled: attempts to balance the differences between areas with relatively little vs large amounts of travel

#### 2.1 Relating road deaths to motorization

Since Smeed's influential paper was published in 1949, there has been much work and controversy on the relationship between the road death rate—as measured per head of population, or alternatively per registered vehicle—and the motorization level—the number of registered vehicles per head of population. The comparison of road death rates between different countries was a large area of interest in the following years. In particular, work was done on comparing the road death rates between developing and developed countries. See Haight (1980), Jacobs & Cutting (1986), Jacobs & Hards (1977), Mekky (1985), Smeed & Jeffcoate (1970) and Wintemute (1985) for example.

#### 2.2 Smeed's equation

Smeed (1949) obtained the following well-known equation relating D the number of deaths due to road accidents,  $V_1$  the number of vehicles, and P, the population, using least squares analysis:

$$D/P = 0.0003 (V/P)^{1/3}$$

and then derived, via simple algebraic manipulation,

$$D/V = 0.0003(V/P)^{-2/3}$$

According to Weiss (1985), Smeed's law of traffic safety, which was derived from a set of 1938 data, was a result that is 'still cited and that is apparently still valid.' He found that it still 'gives a satisfactory fit' to more recent data from 1980, as shown by Adams (1985). Smeed's 1949 paper has had considerable impact in the field of road safety, especially since it was one of the earliest attempts to quantify the relationships between road fatalities and easily obtained macro-data. We have found numerous references to his work in the road safety literature. Not only have researchers applied his model to their own data, but they have also used his model as a basis for more complex models incorporating the effects of population size and the number of vehicles for different countries and regions within countries. See Hampson (1982), and Preston (1982).

However, in the last decade, the accuracy of Smeed's formula has been questioned by many, including Haight (1980), Jacobs & Sayer (1983), Wintemute (1984), Mekky (1985) and Andreassen (1985). In particular, Andreassen (1985) conducts a thorough review of the theoretical basis for Smeed's formula and finds many problems:

- The original regression is different in form from what is known as Smeed's formula. To obtain the formula, the equation containing the coefficients estimated by least squares was simply *algebraically* manipulated. This rendered suspect both the values and the accuracy of the coefficients of the new formula.<sup>1</sup>
- The equivalence of the different forms of Smeed's formula was a coincidence, due to the data set used and rounding performed.
- The formula encouraged people to assume that P and V account for all the variation in D.

Andreassen (1985) then proceeded to show that:

- The values in Smeed's formula were not applicable to many different countries.
- Deaths per vehicle was not a good basis for international comparisons.

Despite the fact that the coefficients derived in Smeed's formula are not always applicable, Weiss (1985) reminded us that 'his work is important in introducing a way of thinking about traffic problems on a large scale.' Minter (1987) echoed Andreassen (1985) in suggesting that separate constants might be required for different countries, but commented that 'Smeed's formula still gives remarkably good prediction of accident rates over a wide range of conditions.'

In a later paper, Andreassen (1991) further investigated the problems with the application of Smeed's 'formula' to various data. In particular, the problem of spurious correlations between independent and dependent variables may arise when one variable is used to calculate both the independent and dependent variables.

Another problem with Smeed's model was that the measure of goodness-offit was not significant at the usual levels.

ì

<sup>&</sup>lt;sup>1</sup>Note that when these approximations were made, the advanced computing regression tools of today were not available to allow simple changes of the regressed variable.

### 2.3 Motorization level

Mekky (1985) defined the motorization rate m as follows:

$$m = \left[ \left\{ \frac{m_{\infty}}{m_0} \right\}^{1/n} - 1^2 \right] * 100$$

where  $m_{\infty}$  is the final motorization rate,  $m_0$  is the initial motorization rate, n is the number of years during which the change in motorization level took place.

Then the number of road fatalities of several countries in a given year was given by a Smeed-like formula:

$$\frac{F}{P} = a(\frac{V}{P})^{1-b}$$

where V is the number of vehicles, F is the annual fatality rate, P is population.

The elasticity, b, of the annual fatality rates of several countries, F, was regressed on the motorization rate:

$$b = \alpha + \beta m$$

The regression was highly significant for rich developing countries, and moderately significant for industrialised countries during the fifties. Thus, the original hypothesis that rich developing countries experience worse road fatality rates than already developed countries was not supported by this evidence.

#### 2.4 Exposure

A number of in-depth studies have considered the probability of a fatality given the amount of exposure a person has had on the roads. It seems reasonable that if a person only travels once a week, they are less likely to be exposed to dangerous situations, than someone who travels every day. On the other hand, someone who travels regularly should have a better technique than someone who travels irregularly!

Jovanis & Chang (1989) wrote:

'A study of accident occurrence alone is generally not sufficient to obtain a complete understanding of accident risk. This is because the occurrence of accidents, as reflected in accident reports, for example, must be compared to the number of opportunities available to be involved in an accident.'

The amount of exposure could be measured by the hourly traffic volume, as done by Oppe (1979), Ivey et al. (1981), Ceder & Livneh (1982); average daily traffic volume and total VMT as described by Jovanis & Chang (1986). Other measures used include: travel speed as done by Hall & Dickinson (1974), Lavette (1977); tonne-miles; passenger-miles; vehicle registrations; weather and vehicle type from Jovanis & Delleur (1983); sales of gasoline, eg Fridstrøm & Ingebrigtsen (1991); and population.

One of the main problems identified was how to combine exposure, which is aggregated on a daily, weekly, monthly or most often yearly basis, with accident data, which is discrete by nature.

Johnson & Garwood (1971) offered the following opinions on the advantages and disadvantages of many different measures of exposure.

Measure	When appropri- ate	Comment
per head of pop- ulation	pedestrian data	vehicles contain one or more oc- cupants
per occupant km	effect of age im- portant	exposure levels per head of popu- lation varies more with age than it does for pedestrians
per km road	indicates ideal location for re- medial measures	
per vehicle kilo- metres travelled	compares differ- ent types of road	allows for traffic flow

Jovanis & Chang (1986) included automobile and truck VMT derived from a toll collection system, the weather as measured by average hours of snow and rain for the toll road, and a weekend binary variable as explanatory variables. They retained all but the weekend variable as valid coefficients in the resulting loglinear regression.

See the section on Problems with Regression Models for a discussion of the statistical problems with the use of simple ratios to represent risk.

#### 2.5 Risk compensation

The paper by Fridstrøm & Ingebrigtsen (1991) considered two basic types of risk compensation. The first theory, advocated by Peltzmann (1975) and Andreassen (1991), is that the relative accident risk to road users who are not benefited by the introduction of a particular legislation or countermeasure may increase, yet result in no change in the overall risk to road users of all types. Thus the response variables need to be disaggregated by road user type to obtain a clearer picture of the effects of the explanatory variables on different road users.

Alternatively, drivers may change their behaviour in response to changes in their environment, including attempts to increase road safety, such as changes in legislation. This change in behaviour may mean that drivers are less careful in a safer environment, and thus make the roads less safe for pedestrians and other non-occupants of motor vehicles. Since it was difficult to measure these changes in behaviour, an indirect method to account for them was used. It involved the simultaneous regression of the number of fatal crashes and the number of fatal casualties per accident (also referred to as the gravity index) on the same group of explanatory factors.

The authors' results were not conclusive. The proportion of new drivers was the only explanatory variable whose effect was opposite for the number of crashes and the gravity of crashes. The effects of snowfall, road improvement and wine consumption were opposing for the two response variables, but not significant in explaining the gravity of crashes.

The model form used by the authors is discussed in more detail in section 8.3 and a summary of the results for various explanatory variables can be found in section 6.

#### 2.6 Hazard models

In a later paper, Jovanis & Chang (1989) used a hazard model to incorporate this idea of risk (or differing amount of exposure.)

There are a number of different hazard models to choose from:

- Independent competing risk model Under this model, the failure of one risk component would cause a traffic accident. However, it is widely accepted that most traffic accidents have many contributing factors. See Treat et al (1977) for example. Also, it is hard to precisely define the failure of a risk component such as weather and road conditions.
- Accumulative hazard model Each individual risk component contributes some hazard to the system depending on its level. Hazards accumulate and the system fails when the cumulative hazard level reaches a threshold value. This model would be appropriate when the levels of all the risk components may be easily measured; and each individual risk component does have some effect on system failure.
- Latent system hazard model Here the probability of a failure at time t is determined by the total hazard contributed by the level of each risk component at that particular time. This overcomes the problem of defining a 'failure' for risk components, and also of specifying the specific causes of accidents.

The population used was a fleet of trucks in the US, since the levels of the various risk components were adequately recorded. They modelled the probability that a vehicle survives until time t in both an additive and a multiplicative latent hazard model.

The individual risk components included were: winter, night, age, weight & experience of the driver, hours driving in the last 8 days, and recent hours off. They found that the risk of a severe accident is strongly related to driver

experience and environmental factors (winter, night), whereas the risk of minor accidents is related most strongly to the number of hours driving.

Chang & Jovanis (1990) extended their hazard model further to account for a number of trials.

The authors noted that since accidents are such rare occurrences, an enriched sampling of failure vs censored data would be required. For instance, the authors mention that Chang (1987), in a study on truck accidents, required 6000 nonaccident trips to obtain a risk factor that significantly contributed to accident occurrence.

#### 2.7 Induced exposure Index

Janke (1991) showed how exposure should not be represented by a simple number of accidents to VMT ratio, which may inflate the exposure risk of people who generally are only involved in short trips. Short trips usually remain within the city and residential areas, involving more stop/starts and much denser traffic. Longer trips comprise mostly freeway travel, which has been shown to have a lower rate of accident incidence than nonfreeway travel. Janke cited evidence from the California Bus, Transportation and Housing Agency.

Furthermore, a linear relationship between mileage and the number of accidents could also inflate the risk.

Cerrelli's induced exposure method (1973) used a hazard index:

hazard index = 
$$\frac{\text{liability}}{\text{exposure}}$$

where

exposure index =  $\frac{\% \text{ mnocently accident-involved drivers in a category}}{\% \text{ licensed drivers in a category}}$ 

and

liability index = 
$$\frac{\% \text{ accident responsible drivers in a category}}{\% \text{ licensed drivers in a category}}$$

The main problem with this method was that assigning responsibility for the accident could be highly subjective, and vary widely between different observers, both within the same region, and from different regions. Wasielewski & Evans (1985) created an induced responsibility model similar to Cerrelli's induced exposure model, but relaxed the assumption that only one driver is responsible. They assigned different degrees of responsibility to crash-involved drivers. Nevertheless, it remained a highly subjective measure.

¥.

# Part II Choice of explanatory variables

## 3 Before-and-after studies

These papers are of interest to our investigation because the models used often can be applied in an explanatory context. Over the last decade, the techniques for comparing road fatalities before and after some countermeasure has been introduced, have become more sophisticated.

#### 3.1 Simple before/after comparison method

The simplest, most naive method, consists of fitting a simple regression line over time to the data, and determining whether the slope is different before and after. (See Brinkman (1986).) Recently, however, more explanatory variables have been incorporated into the regression before determining whether there has been some change. Wagenaar (1984), for example, suggested that economic factors may overwhelm the effects of legislative countermeasures, as supported by common health literature findings that 'negative changes in economic conditions, such as increasing unemployment rate, are associated with increased incidence of health problems.'

Persaud (1986) believes that many traffic engineers hold the misled belief that the effectiveness (or reduction in accidents) of a safety measure depends on the number of accidents at the location before installation of the countermeasure, and that the established effectiveness of a safety device may often be attributed to the 'regression-to-mean' effect.

This 'regression-to-mean' effect is such that it is entirely due to chance that high values may be observed before a change, and low values afterward. A good explanation of this effect can be found in Hauer (1986), p3. Given a random variable which fluctuates around its mean, the best estimate of a future value of this random variable is just its mean. So even if we observe an unusually high value at a particular time, then we still expect the following observation to be the mean. Hence, a downward trend towards the mean is observed after an unusually high value is observed.

A paper by Brinkman (1986) outlined the dangers inherent in using these simplified before/after evaluation techniques. He advocated the use of some basic statistical principles: using data over as long a time period before and after, as is possible; randomised control groups; comparison groups (matching); time series designs; empirical bayesian statistics.

An earlier paper by Johnson & Garwood (1971) suggested that there should be 'allowance for growth of traffic, weather, changes in legislation' and that 'the seasons be equally represented.'

#### 3.2 More sophisticated techniques

Fortenberry et al. (1985) described a 'nonparametric quasi-experimental' method for evaluating countermeasures. The response variable was observed for two different locations: a treatment (with intervention), and a control. They partitioned the observations before and after the intervention as the baseline and operational period respectively. Then a regression line was fitted to the baseline and operational period separately for each location. They proceeded to compare the amount of deviation from the regression line for the operational period of each location using a form of the Wilcoxon statistic. Nevertheless, the statistic they derived was still based on a simple regression over time.

The Expert Group on Road Safety (1978) used contingency table analysis to compare before and after accident rates. They also used the rest of Australia as a control when considering data for Victoria, which was the first state to introduce compulsory seat-belt wearing. However, the authors decided that the responses from the other states were confounded by the effect that the widespread publicity of Victoria's new legislation obtained.

Danielsson (1986) proposed that the effect of countermeasures may be overinflated if the only sites where the number of accidents were measured were those notorious for having high accident levels. The author based his method on that used by Hauer (1980a). The number of accidents before the countermeasure was modelled as a Poisson distribution with parameter  $\lambda_i$  for the *i*th geographic location affected by the countermeasure. The number of accidents occurring after the countermeasure was also modelled as a Poisson distribution with parameter  $\alpha \lambda_i$ , where  $\alpha$  represented the proportional change in the accident rate.

Danielsson compared the performance of three estimators of the number of accidents after the countermeasure was introduced. The traditional estimator lost 30% efficiency; the maximum likelihood estimator was very accurate, independent of the true value of  $\alpha$  and also best for large values of  $\alpha$ ; and Hauer's estimator underestimated the effect of the countermeasure, although it performed well for small  $\alpha$ .

## 3.3 Arguments against the effectiveness of countermeasures

Several authors, especially Peltzmann (1975) and Minter (1987), have suggested that the fatality rate is decreasing irrespective of any changes in legislation.

## 4 Regression analysis

In the road safety literature, it has become more common for researchers to characterise studies by the way in which the variables of interest have been observed. This gives rise to three 'types of studies':

- longitudinal studies (often referred to as time-series studies), where the observations are obtained from regularly spaced time periods, measured at the same location. Here the emphasis of the study is how the variable is changing over time.
- crossectional studies, where the observations for particular time periods are gathered from different geographical locations. Here the emphasis is on determining whether the observed phenomenon recurs in different locations.
- pooled crosssectional longitudinal studies which intend to look at the effect of both the geography and time on the variable (hence the term spaciotemporal.)

Zlatoper (1989) presented a 'selective survey' of the United States literature on road crash fatality rates. He began with a description of a pivotal study by Peltzmann (1975). Its controversial finding, that there was no decrease in road fatalities after the implementation of new legislation, sparked many criticisms. Peltzmann's study is an ideal example of the typical regression analysis of road crash fatalities, and the subsequent criticisms illustrate the many finer points which need to be considered.

This paper is also of interest because it supports the unpopular view that road fatalities are decreasing anyway, irrespective of new legislation. See the section on Before-and-After analysis for more discussion on this long term downward trend in fatalities

#### 4.1 Peltzmann's initial model

Initially, the response variable in Peltzmann's (1975) model was the annual fatality rate per vehicle mile, standardised for type of driving (urban, rural) and for the type of road (highway or other). Three different road crash fatality rates, also called motor-vehicle death rates, in the US were modelled: total, non-occupant (pedestrians, bicyclists, motorcyclists) and vehicle-occupant (total minus non-occupant). The study compared annual data from prelegislation years, 1947-1965 to postlegislation years, 1966-1972.

The explanatory variables used were as follows:

Cost of an accident In order to measure the cost component of an accident that is typically insured, the author used the index of direct accident costs (property damage and medical costs) multiplied by an insurance loading factor (ratio of premiums to benefit paid.)

Income The real earned income per adult over fifteen years of age.

- Alcohol The alcoholic intoxication level amongst the population at risk was measured by the consumption of distilled spirits per persons over fifteen years old.
- Driving speed was measured by the estimated speed of motor vehicles on noninterstate roads during offpeak hours.
- Youth The driver age distribution was represented by the ratio of 15- to 25year-olds in the population to older people.

#### Trend

#### 4.2 Criticisms

**Correlation between explanatory variables** Joksch (1976) found that the income, time trend, and speed variables were highly correlated. Robertson (1977) noted that the paired correlations between the explanatory variables differed in the prelegislation and postlegislation periods. Peltzmann (1976) used first differences to avoid these correlations.

Other methods used to remedy multicollinearity are: obtaining more data, identifying relationships between the explanatory variables, and omitting one of the correlated variables.

Choice of response variable: rates or raw totals? Joksch (1976) echoed many critics of Smeed (1949) in questioning the validity of the use of a fatality rate as a regression response variable. Basically, using the ratio of two variables as a response variable assumes a linear relationship between the two. If this assumption is wrong, then additional correlation may wrongly be introduced into the model.

In reply, Peltzmann (1976) investigated two things. First, the number of fatalities was regressed on vehicle miles. A regression coefficient which was insignificantly slightly less than one was reported. However, it was noted by Robertson (1977) that this coefficient was not significantly different from zero either.

Secondly, the number of fatalities was used as the response variable in the initial model. This produced a similar pattern in the regression coefficients as obtained when the fatality rate was used a response variable.

L

Unstable regression Joksch (1976) also found that the regression was unstable after performing a validation analysis. The coefficients of some regressors were not consistently significant under a slight change to the model, for example, addition or deletion of other regressors, or a change in the functional form.

Omitted variables The major criticism was that various important variables had been omitted from the regression. Thus, the change (or lack thereof) in fatalities, could not solely be attributed to the introduction in legislation, since Peltzmann's initial model only accounted for six other factors. Furthermore, the estimated effects of the factors that are included in a model could be biased if important variables are not accounted for. Other variables suggested were:

Graham & Garber (1984)	vehicle size distribution, no-fault in-
	surance
Joksch (1976)	highway improvements, weight and size of vehicles
Zlatoper (1984)	volume of driving, vehicle size, type of driving as measured by the ratio of ru-
	ral to urban vehicle miles
Garbacz (1985)	miles of interstate highways

Correct measurement of the desired variables Robertson (1977) questioned whether the cost-of-an-accident variable was a valid measure of crash costs, and suggested alternative expressions for a few of the other variables. The alternative measurement for 'youth' of the driving population was the ratio of drivers in the 15-24 year old age group to the total number of drivers involved in accidents The percentage of motorcycles registered as compared to all vehicles registered could account for the shift in risk from occupants to non-occupants. The alcohol measure could account for beer as well as distilled spirit consumption

The author found that the correlation problem was no longer present, and that the impact of legislation was significant when the definitions of these variables were used. Peltzmann commented that data-dredging will often produce the required results, and noted the arbitrary nature of the definitions, questioning their theoretical justification.

**Choosing the correct functional form** Graham & Garber (1984) found that the model was sensitive to the functional form used. They suggested that a logarithmic form may be incorrect, and that the linear form should at least be investigated.

Identifying hierarchy of relationships amongst explanatory variables Instead of using just one regression equation relating all of the explanatory variables, a number of authors have suggested the use of simultaneous regression equations. (See Zlatoper (1989) and Hakim *et al* (1991) for more details.) Some explanatory variables could be regressed on others to obtain estimates of these variables to be used in the fatality regression. This also ensured that prediction could proceed more smoothly, since predicted values of the regressors could be used in the main regression equation.

### 4.3 Crosssectional studies

Even though the emphasis in these studies is not the pattern in fatalities (or fatality rates) over time, they still offer interesting suggestions for the choice of explanatory variables.

Peltzmann's (1975) model used the following variables to explain fatality rates for 1970 in different states of the US. The per capita death rate, adjusted for the effects of interstate highway travel was related to:

- the fraction of all cars, of which had been subjected to the new motor vehicle regulations, or younger
- per capita fuel consumption
- speed limit on main rural roads
- ratio of urban to rural driving
- vehicle mile in urban and rural context
- alcohol, youth, accident cost variables as described above
- disposable income per capita
- ratio of earned income per adult to unearned income per capita

L

#### 4.4 Problems with regression models

Hautzinger (1986) outlined a number of points to be careful of when performing a linear regression:

- A high value of  $R^2$  does not necessarily indicate a good model.
- A sound theoretical basis is required for the structure of the model and the choice of the explanatory variables.
- Multicollinearity between variables should be checked.
- Autocorrelation of errors, and instability of error variance should also be checked.
- High level aggregation should be avoided wherever possible as this reduces the sample size, often to less than twenty.
- Standard linear regression models should not be applied to data disaggregated by geographical location; instead multiregression or temporal crosssection models should be used since they allow for decomposition of error terms into components. Otherwise, estimation, inference and forecasting may become inefficient.
- In longitudinal studies over time, the significance of terms must be maintained even for different base periods.

Some other criticisms of models, presented by Mahalel (1986) and Andreassen (1991) among others:

- The risk to an individual driver or passenger is often characterised by a simple number of accidents to the amount of exposure ratio. However, this requires a linear relationship between the numerator and denominator, which may cause some problems if risk varies with exposure level.
- Some systems function more effectively at certain levels of exposure and less effectively at others. For example, traffic light reduce accidents in high traffic volume areas but often increase the number of accidents in low traffic volume areas.
- Spurious correlations may be introduced into the model if any two variables are related to a third variable.
- The number of fatal crashes should be used in preference to the number of fatalities, since the latter is a function of the former, and most of the explanatory variables used by researchers actually affect the number of fatal crashes, not the number of fatalities.
- The number of crashes should be disaggregated by road user type. Pedestrian and motor/pedal cyclist deaths have been shown to behave differently to motor vehicle deaths.

## 5 Economic relationships

The state of the economy is known to affect many different phenomena. A number of researchers have included economic variables in explanatory models for road safety.

Foldvary et al (1971) found a relationship between economic trends and road accident trends in the Australian context. He stated that 'slump conditions discouraged inexperienced drivers from entering the driving population, and these were the drivers with the worst driving record...the total milage travelled during a slump would be less.' So, he suggested that an ideal covariate for measuring the extent of a slump would be the number of teenagers becoming eligible to drive in a particular year. However, as this was difficult to obtain in practice, Foldvary used the urbanisation rate, and the number of car registrations, to explain the number of fatalities in a given year. Here, the urbanisation rate was simply the percentage of the population which reside in urban areas. He found that the log of the log of the fatality rate (per vehicle) varied linearly with the percentage of urbanisation.

Aldman (1980) showed that 'a curve representing business cycle variations also fits the accident curve quite well', when looking at monthly data.

Hedlund et al (1984) compiled results from other studies and FARS data for 1980 and 1982 to explain a 14% decrease in fatalities in the US. The authors concluded that economic effects, reflected by VMT, were the main contributing factors, with driver-education, increased RDU and decrease in number of youths having little effect.

#### 5.1 Recession impact

A paper by Eshler (1977) related the state of the economy to the fatal accident rate in the US. The work of Joksch & Wuerdemann (1973) was cited, among others, as an example where a strong relationship between the economy and fatalities was found.

The general economic measure considered was the delineation of recession and non-recession periods as defined by the National Bureau of Economic Research (NBER). Briefly, a recession period may be indicated by:

- duration a decrease in the real Gross National Product (GNP) for two consecutive quarters, and a decrease in industrial production over six months
- depth a 1.5% decrease in real GNP, a 1.5% decrease in nonagricultural employment, and a 2 point rise in unemployment to a level of at least 6%
- diffusion a decrease in nonagricultural employment in more than 75% of industries as measured over six monthly periods, observed for six months or longer

When graphed against the annual fatality rate, it was noticed that the recession periods coincided with a 'levelling out' of the fatality rate. The author also noted that the energy crisis of 1974 and the lowering of the speed limit at approximately the same time would have affected the fatality rate.

The unemployment rate and the average number of hours worked were chosen as more direct measures of the impact of the economy on drivers.

The resulting model was fitted in steps, but was equivalent to fitting a regression model with a nonlinear long-term trend, which accounted for 89% of the variation; a linear unemployment term, which was significant; and the average number of hours worked, which wasn't significant.

#### 5.2 Indirect economic effects

Many authors have found that the energy crisis in 1974 affected many different variables which may have affected road safety. For instance, Godwin (1984) found that gasoline prices increased, reduced speed limits legislation was introduced, and even patriotic fervour may have altered the behaviour of drivers.

#### 5.3 Unemployment

Cooper (1984) and Partyka (1984) found that unemployment was a strong predictor for the frequency of accidents. Mercer (1985) took this one step further and incorporated several other variables as well as unemployment into a model for the number of fatalities. The predictor variables used were:

- monthly unemployment figures: for the percentage males and females, aged 15-24 and over 24.
- monthly average percentage of passenger vehicle casualty accidents that were alcohol-related. This was claimed to be more stable and therefore a superior predictor than just the prevalence of Driving under the Influence (DUI).
- monthly average monthly percentage of occupants in passenger vehicle casualty accidents who were using restraint devices. This was claimed to be underestimating the actual rate of restraint device usage (RDU) by 12-15%.
- monthly average age of drivers in passenger vehicle casualty accidents. This average reflected the changes in age of the driving population as a whole, but was considered to be of questionable validity.
- monthly average percentage males as drivers in passenger vehicle casualty accidents.
- time (linear variable)

Mercer then calculated Pearson's correlation coefficient for each of the relationships among the predictor variables and the response variable (the number of fatalities or the number of casualties.) Also calculated were the partial correlation matrices, designed to control for the effects of the unemployment variables and for the effects of unemployment and driver demographics.

The findings were that as unemployment rose, the average age of drivers in casualty accidents rose, and the percentage of males as drivers fell. This was easily explained by hypothesising that unemployment increases removed young male drivers from the driving population. The same change in driver demographics was observed as RDU and DUI rose.

Mercer's conclusions were that

"... changes in unemployment levels arguably produce changes in driver demographics, which then appear to be related more strongly to changes in accident frequency and severity than are changes in drinking driving and restraint device use ... changes in traffic accident figures must be considered within the context of economic trends and driver demographics in addition to driver-related behaviour such as restraint device use and drinking driving."

Wagenaar (1984) suggested that there seem to be two different ways in which unemployment may concervably affect road safety:

- 1. High unemployment and the associated reduction in disposable income could lead to decreased travel by private vehicle, and thus reduce the number of road crashes, since the exposure to risk is reduced.
- 2. Alternatively, high unemployment could lead to more stress in the driving population, which could in turn cause more aggressive driving, leading to an increase in road crash rates.

Wagenaar aimed to establish which of the above hypotheses were supported by monthly data from January 1972 to January 1982 in Michigan, on the number of drivers injured and killed in road crashes. A Box-Jenkins time-series model was employed; details and conclusions are presented in the section on Time Series analysis.

Partyka (1984) regressed the annual number of fatalities in the US on the number of unemployed workers, the number of employed workers, the size of the the nonlabor workforce with an intervention variable for the 1974 fuel crisis and the lowering of the US national speed limit.

The value of  $R^2$  for this model was very high, 98%. Nevertheless, Partyka warned us of the following limitations of the model:

1. The effect of omitted variables was not predictable. Some variables that were omitted, such as improvements in roadway and vehicle design, and driving habits, were difficult to distinguish from the long-term trend.

- 2. The model only showed where a relationship existed between variables, and did not imply any cause and effect relationship.
- 3. Predictions beyond the range of the data would be unreliable.
- 4. The model assumed independent errors.

In a follow-up paper, Partyka (1991) explored how the extra data from 1983– 1989 affected previous results. The value of  $R^2$  was virtually the same; and the coefficient estimates varied only slightly and were still significant at the 1% level.

However, the actual fatalities for 1983-1989 were very different to those predicted by the model based on the 1960-1982 data. Omitting the variable for the number of employed workers gave slightly better results, but the predictive power of the model was still not good enough. The additional data suggested that 'something new' happened in 1983, which could possibly be explained by the increase in seatbelt use and the decrease in DUI.

Reinfurt et al (1991) also extended Partyka's 1984 model in several ways. They considered suicides and homocides as well. They stratified the fatality data by age (16-24, 25-44, 45-64 and 65+), race (white vs non-white), and sex. Two types of models were investigated: regression models based on Partyka's 1984 model; and structural time series models based on Harvey and Durbin (1986).

The authors reported parameter estimates and standard errors, and the value of  $R^2$  for a full regression model including Partyka's unemployment variables. These values were not reported for a model which did not include the non-significant variables. They found that the best fits were obtained  $(0.70 < R^2 < 0.95)$  for the models for the youngest and oldest agegroups, for both races and both genders.

They found that an ARIMA model based on Harvey and Durbin's (1986) structural time series model was better at predicting short-term variation. Two intervention variables, for the oil embargo and the introduction of the 55mph speed limit, had a significant effect on fatalities. The only unemployment variable included in this model was the level of employment.

It was interesting to note that the unemployment variables contributed significantly to motor vehicle deaths, but not to suicides and homocides.

#### 5.4 Availability of medical facilities

The definition of a road fatality differs from country to country. It can range from death at the scene of the accident (in Spain) to death within 30 days of the accident (in Australia.) In the intervening period between the accident and death, the casualty is highly likely to be hospitalised, or at least examined by a physician. Thus, the availability of medical facilities could affect whether a serious injury becomes a fatality. Jacobs & Hards (1977) found that fatality rates for several countries, including the USA, were correlated with the level of medical facilities available, expressed in terms of population per physician and population per hospital bed. This model failed for a few developing countries.

The findings were as follows. As the number of vehicles or the vehicular density increased, the fatality and casualty rates decreased. As the population per hospital bed increased or the per capita Gross National Product (GNP) decreased, the fatality rate increased. The fatality index was most significantly affected by the population per physician, and to a lesser extent by the number of vehicles and the per capita GNP.

Jacobs & Cutting (1986) extended Smeed's (1949) paper. For several different countries, for three separate years, they modelled log fatality and casualty rates per registered vehicle, and the log fatality index, the gravity index, the proportion of all persons injured who are killed. The explanatory variables considered were: vehicles per km road, road density (km road per km<sup>2</sup>), vehicles per person, GNP per capita, population per hospital bed, and population per physician. Three different response variables were investigated: the log fatality rate (per million vehicles), the log casualty rate (per million vehicles) and the log fatality index.

Vehicles per person were highly significant contributors to the increase in the three different response accident indicators. In addition, GNP per capita, vehicle density, and population per hospital bed, were found to be significant contributors to the fatality rate. Population per hospital bed was also the most significant contributing factor to the increase in fatality index.

Lave (1985) found a negative, sometimes significant, effect for access to emergency medical facilities on road fatalities, when accounting for other variables. See section 6.3 for more details on the model used.

#### 5.5 Investment into traffic safety facilities

The model used by Murata (1989) accounted for the stock of traffic safety facilities, the annual budget normalised by the Gross National Product (GNP) and the total trip length (VMT).

$$S_N = S_0 + \sum_{t=0}^N B_t \exp{-\lambda t}$$

where  $S_N = \text{stock}$  of traffic safety facilities after N years from the initial year;  $S_0 = \text{stock}$  existing in the initial year;  $B_t = \text{yearly budget}$ ;  $\lambda$  is a parameter to be estimated. Both stock and budget were expressed as the proportion to the GNP. Then, to relate this relationship to the number of accidents every year,

$$\frac{E_t}{T_t} = \frac{k}{S_t}$$

where t was the year;  $E_t$  was the estimated number of accidents;  $T_t$  was the total trip length (VMT);  $S_t$  was the safety stock as estimated from the equation above; and k was a constant.

Hence,

$$E_t = kT_t / \left\{ S_0 + \sum_{i=t_0}^t B_i \exp -\lambda(i-t_0) \right\}$$

where  $t_0$  was the initial year, and k then represented the effectiveness of the traffic safety facilities in prevention of accidents.

Frigstrøm & Ingebrigtsen (1991) calculated monthly indices describing the relative increase in road capital by country and national authorities per km road length in Norwegian countries. Investments in county roads increased safety, but decreased for nationally-controlled roads. The authors did not expect to obtain this type of result, and suggested that classification error between county and national roads may have influenced the results.

#### 5.6 Socioeconomic variables

#### 5.6.1 GNP

Road fatalities are generally expected to rise as GNP increases according to Havard (1979) and Haight (1980), for example. Wintemute (1985) used an epidemiological argument to explain this. At specific levels of economic development, nations experience: a demographic transition, when infant mortality rates and fertility rates decrease; and an epidemiological transition, when infectious and nutritional diseases are less common than chronic, degenerative conditions. Thus, as nations achieve this level of socioeconomic development, their use of motor vehicles increases.

Since the per capita GNP is a 'widely relied upon indicator of development', appropriate indicators of the socioeconomic development level were taken to be: the per capita gross national product (PCGNP); income distribution, which could be measured by the Gini indicator; and the population. The Gini indicator is a number between 0 and 1, where higher values indicate a large deviation from a uniform income distribution.

A direct but weak correlation was found between economic development, as measured by the PCGNP and motor-vehicle mortality. The relationship was strongest for low development (low PCGNP). Thus poor and intermediate countries had rapidly increasing motor vehicle mortality rates. It was also found that for the poorer countries, the income indicator accounted for much of the variation in fatality rates. Winternute suggested that the income distribution may be related to the level of urbanisation.

The author cited Jacobs & Sayer (1983) who suggested that the model may be extended by considering geography, climate, level of urbanisation, traffic mix and flow, infrastructure and development, availability of medical services, and cultural trends.

#### 5.6.2 New car registrations

Hautzinger (1986) looked at how the general socio-economic climate could affect the annual rate of accidents amongst insured people covered by a large insurance company. It was considered appropriate to use general population measures since the company's marketshare was relatively large. The authors considered a number of socio-economic measures:

- total annual car mileage
- aggregate income variates
- the number of new car registrations
- consumption of private households
- price indices
- labor force data

Finally, two models were chosen. The first modelled the accident rate in year t as a loglinear regression model incorporating the effects of time and the seasonal effect of the time series of car registrations (ratio of the number of new car registrations in year t to the corresponding trend value.)

Two different base periods for estimating the regression coefficients were used: 1961-1982 and 1970-1982. The coefficient for the time trend was significant in both cases, but the coefficient for the seasonal effect of new car registrations was significant only for the shorter base period. The  $R^2$  value decreased from 0.84 to 0.54 as the base period was shortened. The Durbin-Watson statistic for the shorter base period was significant, indicating the possible existence of autocorrelation. This prompted the use of growth rates.

The growth rate of the accident rate was regressed on the growth rate of the seasonal effect of new car registrations. (The growth rate of  $Y_t$  is  $\Delta Y_t = (Y_t - Y_{t-1})/Y_{t-1}$ .) An  $R^2$  value of 0.17 was obtained for the longer base period. However, the coefficient for the change in new motor vehicle registrations was significant over both base periods. Furthermore, the Durbin-Watson statistic indicated that the autocorrelation had been reasonably accounted for.

This model for accident rates was then used in a supermodel for accident damage costs.

Fridstrøm & Ingebrigtsen (1991) found that the proportion of new drivers adversely affected safety, yet unexpectedly favourably affected the severity of crashes. This could be due to the reduced speeds of inexperienced drivers, which would lead to decreased severity of crashes yet possibly cause more crashes. However, previous studies, such as Wasielewski (1984), found that speed was inversely related to age and/or experience.

Ł

### 5.7 Discontent

Sivak (1983) found evidence to support the hypothesis that as violence and aggressiveness in society rise, the number of injuries in road crashes increases. The author measured the level of violence by the number of violent and property crimes, the number of police calls for domestic disputes, suicide rates, and the number of worker strikes.

### 5.8 Youth

Wagenaar (1983) finds that a higher minimum legal drinking age is associated with lower fatalities amongst young drivers in various US states.

## 6 Other explanatory variables

Apart from economic considerations there are a number of other explanatory variables which have been investigated by various researchers.

#### 6.1 Level of vehicle inspection

A number of studies have considered whether the level of vehicle inspection has affected the numbers of road fatalities, in varying degrees of sophistication. One of the first studies in this area, Mayer & Hoult (1963), looked at whether four different categories of inspection in different states affected fatalities over a period of twelve years. Buxbaum & Colton (1966) based their analysis on Mayer & Hoult, but included extra variables, namely gasoline consumption per vehicle, and the number of vehicles.

Fuchs & Leveson (1967) included many more econometric variables: age of driver, education, median income, fuel consumption per capita, population density, alcohol consumption per capita, socio-economic variables and a binary variable for inspection. The result was a non-significant contribution from the level of inspection.

A more recent study by Loeb & Gilad (1984) used an even more complex regression on among others, the annual number of deaths, and the death rate per VMT: time variable as an indicator of technological change, maximum highway speed, gasoline consumption, the number of licences revoked for DUI, per capita personal income, population, the number of motor vehicle registrations, the number of licenced drivers, vehicle mileage GNP price deflator, inspection level, dummy variable for World War II and the great depression. They found that time, personal income, population, level of inspection and World War II made the most significant contributions to the regression.

Garbacz & Kelly (1987) investigated the impact of compulsory motor vehicle inspection with the inclusion of three explanatory variables representing three levels of inspection: biannual, annual and spot. The basic model was taken from Garbacz (1985), where a double-log functional form was used for all of the variables. Fatality rates for three categories of road user were modelled: total, occupant and non-occupant. The explanatory variables used were: the real disposable income per driver; weighted sum of the medical care and auto repair services indexes of the CPI; Peltzmann's youth indicator; per capita consumption of spirits, wine and beer adjusted for alcohol content; ratio of vehicles equipped with regulated safety equipment; dummy variable for 55mph speed limit legislation; miles of interstate highways.

Five models were investigated: three separate models for each inspection level, one model including all inspection level variables, and one model including the ratio of the vehicles subjected to any inspection level to the total number of registered vehicles.

The authors found that income, alcohol and youth had a positive effect on

the fatality rate, whilst accident-cost, 'safe' vehicles, speed limit legislation, and interstate highway variables had a negative effect. Apart from the alcohol and income variables, all regression coefficients were significant at the 10% level or less. However, inspection level did not appear to affect the fatality rate.

White (1986) used local variables such as the age of the driver and time since vehicle was last inspected to explain fatalities. Conflicting results were obtained.

Fridstrøm & Ingebrigtsen (1991) found that less roadside technical controls were associated with increasing occupant injuries and decreasing nonoccupant injuries. This supported the hypothesis that increased perceived safety of drivers results in lowered safety for other road users.

#### 6.2 Rural and urban speed limits

Godwin (1984) showed that a change in fatality rate varied linearly with a change in the posted rural speed limit.

This was followed with a study by Fieldwick & Brown (1987) which considered the relationship between the number of fatalities and the number of casualties and several regression variables: population, the number of vehicles, and the general urban and rural speed limits. The urban limits were divided into three categories: motorways, roads and others.

The results were as follows. The number of vehicles was highly correlated to population and both of these variables were highly correlated to the number of fatalities and the number of casualties. The three rural speed limits were moderately correlated (from 0.52-0.55). Surprisingly, rural and urban speed limits were almost independent, and both were independent of population and vehicle numbers. The final regression equation, based on the log of fatalities, included coefficients for population, the urban speed limit, and a combined rural speed limit.

#### 6.3 Variability of speeds

Hauer (1971) found that it was the variability in the speeds of different vehicles on the highway which increased the probability of an accident, as opposed to absolute speed. However, as speed increased, driver reaction times decreased and the force of impact increased, also contributing to higher accident risk. Thus Godwin (1984) proposed a model of the *change* in the fatality rate, as a linear function of the *change* in posted speed limits. A significant positive contribution from the change in posted speeds to the change in fatality rates was found.

Lassarre (1986) used a loglinear regression model with an ARIMA error term to model the monthly number of accidents and deaths in France. The model used is described below in section 7.1. The study found that as the variability in speeds rose, so did the number of fatalities, in accordance with a study mentioned earlier, Godwin (1984).

This led to the conclusion that homogeneous speeds increase road safety. The seatbelt wearing variables were found to be insignificant, although the traffic volume index was found to be significant. A large residual was attributed to other factors which had not been taken into account: the improvement of roads and vehicles and emergency hospital services over time.

Another study investigating the effect of the magnitude and the variability of speed of fatality data was Lave (1985). The fatality rate per VMT, for six different types of road, was linearly regressed on: average speed; speed variance (85th percentile minus the average); speeding citations per driver; and access to emergency medical care.

This author also found that average speed had no significant effect on the fatality rate, although speed variance had a positive and significant effect. Hospital access had a negative effect which was significant for only a small number of road types.

Hakim et al (1991) concluded that the independent effects of speed limits and speed variability still needs to be resolved, after citing the literature following Lave's controversial findings

#### 6.4 Public transport fares

It seems sensible to suppose that if the public transport fares were fairly reasonable in comparison to motorization costs, then people would be more likely to use this form of transport in preference to private means. If not, they may have to resort to driving their own car, walking, riding a bicycle, or hitching a ride with a friend. This would then transfer some of the potential car drivers to other types of road users (pedestrians, bicyclists, passengers) and may then affect the overall fatality rate. Alternatively, public transport fares are loosely connected to the economic climate, which may also have an effect on fatalities.

One of the first papers to look at the effect of fares in the British transport system was Oldfield (1977). In another study, Allsop & Turner (1986) modelled the monthly number of casualties, fatal & serious and slight, from January 1978 to April 1983. The explanatory variables used were:

- the real level of London transport (LT) fares (after adjusting for general inflation using the Retail Price Index)
- trend over time (a proxy for all other time-varying effects)
- month of year (seasonal effects)
- lying snow<sup>2</sup> (since this changed from year to year)

<sup>&</sup>lt;sup>2</sup> the number of days in the month when more than a trace of snow was lying in St James' Park at 9am
- real level of British (BR) rail fares (as distinct to LT)
- real level of pump prices of petrol
- number of working days per month on which there was a large scale strike in the BR service
- dummy variable for the change of fare in April 1982 onwards. This fare change was significant: double the previous fare levels.

A log-linear regression was used to incorporate all of these explanatory variables as well as a constant, and a severity and monthly effect, since the effects were deemed multiplicative.

Also considered were lagged variables, interaction between explanatory variables and the severity or month effects, interaction between severity and month effects. The response variables considered were the numbers of casualties for several different types of road user. The findings were as follows:

- 1. LT fare increase in general led to an increase in casualties to cyclists and to occupants of other vehicles and of cars and taxis.
- 2. The general effect of LT fare changes on pedestrian casualties was weak. There was strong evidence of an effect after an increase in the number of casualties after March 1982.
- 3. For users of powered two-wheeled vehicles and occupants of PSVs, there was strong evidence of a change after March 1982 that contrasted effects of fare changes in general.

These findings could be explained in part by two factors: the walking associated with use of public transport was much greater than the extra walking undertaken as an alternative to public transport together with the increase in extra-vehicular traffic; and a shift from public transport to cars and cycles.

Some extensions to the model that were suggested by the authors were allowing for the effect of legislation regarding seatbelts, use of powered two-wheeled vehicles by learners; nonseasonal variation in wet weather; unemployment levels in the Greater London district; and changes in levels of the LT series.

# 6.5 Airline deregulation

Bylow and Savage (1991) looked at the effects of deregulation on the airline model using an econometric 'structural' models based on the assumption of profit maximisation. The explanatory variables they considered were: the total number of airline departures for commuter and jet aircraft; the number of miles of interstate highway; the number of licensed drivers; the average speed of automobile travel; the real per capita GNP; the cost ratio of real air travel price to real gasoline price. The Durbin-Watson statistic indicated that autocorrelation was not a problem. The authors obtained a high value of  $R^2$ , although there was a large degree of multicollinearity between the variables. They claimed that this multicollinearity did not affect the parameter estimates, and so the model could be used for prediction purposes only!

An earlier paper, Evans et al (1990), compared the age and sex profiles of airline passengers to those of the average driver and suggested that airline passengers have a 24.1% lower fatality rate. This was conservative if these people were more predisposed to wearing a seatbelt, refrain from drinking and drive larger motor vehicles.

#### 6.6 Holiday effects

Intuitively, it seems reasonable to suppose that on public holidays there is a marked increase in the number of people travelling all over Australia to meet up with relatives, or to go to tourist destinations. Hence with the increase in people travelling by road, there should be a corresponding increase in the number of road crashes and therefore fatalities.

The paper by Arnold & Cerrelli (1987) showed that there was indeed an increase in the number of road crash fatalities in South Africa during public holidays, particularly over Easter and at Christmas. Ensenberg (1984) modelled the number of accidents for holiday periods and normal times separately, using a log-linear model. He found that the variation from year to year was so great that no long-term trends were discernible.

### 6.7 Using a number of explanatory variables

Thomson (1982) identified various factors which could be contributing to the annual number of fatalities in NSW and Victoria. These factors included:

- vehicle characteristics VKT, % freeways in the road network, vehicle density, private vs bus travel, passengers per vehicle (total population divided by the total number of registered vehicles), age of vehicle stock (three year cumulative new vehicle registrations divided by the total number of vehicles registered), mix of vehicle sizes, rural vs urban travel
- government policy police activity, roadwork investment, quality of hospital services, average speed limits, number of traffic lights installed annually
- driver demographics alcohol consumption, driver age structure, the number of migrant drivers, seat belt usage and motorcycle helmet usage (as determined by surveys conducted by ABS)
- economic variables average household income, annual percent change in real state GDP

ì.

other variables annual rainfall

Several data sources were suggested.

The alcohol consumption could be measured in several different ways. The first is the real per capita spending on beer, wine and spirits deflated by the alcoholic beverages delator, which is provided annually by the ABS. Alternatively, the sales of beer, wine and spirits could be obtained from the Licensing Board. Finally, the estimated absolute alcohol content of alcoholic beverages consumed may also be obtained from the ABS.

Vehicle kilometres travelled could be obtained from surveys of motor vehicle usage, with an interpolation for between survey years based on fuel (petrol and diesel) consumption. The NRMA Data Book is cited as a reference here. This would only be useful for modelling annual data, otherwise the interpolations would be based on a very small amount of data, and thus be unreliable.

Preliminary findings showed that the best model fits for annual numbers of fatalities were obtained for VKT, freeway effects, seat belt usage, rainfall patterns and the mix of vehicle sizes.

#### 6.8 Use of seatbelts

A coverpage story in the Bulletin (Nov 13, 1990) described the introduction of seatbelt legislation as a 'political silver bullet' because of the 'obviously' drastic effect that it had on the number of fatalities from road crashes.

This effect was not restricted to Australia. A Norwegian, Berard-Andersen (1978) investigated the use and effects of seatbelts in twenty one countries and found that 'serious and fatal injuries are reduced by 65-80%'.

Johnson et al. (1980) modelled US monthly fatality data for 1970-79 with changes in VMT, introduction of safety improvements and the implementation of the 55mph law.

Fridstrøm & Ingebrigtsen (1991) found that seatbelt use significantly affected the number of fatal crashes and fatalities amongst motor vehicles, although it did not significantly affect fatalities amongst pedestrians and cyclists, suggesting a lack of risk compensation behaviour.

# 6.9 Blood alcohol concentration (BAC) levels

This topic has been a major focus in highway safety research over the last two decades, with well over 100 papers in the area which have mostly concentrated on local effects models, not macro effects models. In the US, Fell (1982) stated that '... alcohol may be involved in 50-55% of fatal accidents, 18-25% of injury accidents ...' . In fact, in the US, there is an entire publication devoted to the discussion of 'Alcohol, Drugs and Driving.' A paper by Moskowitz et al (1986), published in this journal, is a collection of abstracts and reviews of papers in this area.

The question of whether alcohol consumption per capita is a good index of alcohol-related traffic safety problems was addressed in a study by Mann & Anglin (1988). They considered the impact of the availability of alcohol, indicated by increased hours of sale for example; on-premise availability, places where people would drink and then drive home.

They found a strong relationship between alcohol availability and per capita consumption and the numbers of alcohol related crashes.

In the longitudinal model used by Fridstrøm & Ingebrigtsen (1991), the monthly series of fatal crashes and fatalities significantly increased with increasing convictions *not* due to DUI, yet was not significantly affected by convictions due to DUI.

They also found that the number of accidents was positively correlated with liquor consumption, but negatively correlated with wine consumption. They cited two other technical reports from Canada and New Zealand, which showed similar results and suggested that the age composition of liquor consumers and wine-drinkers was probably different.

Walsh (1987) found that in Ireland, the per capita alcohol consumption increased with increasing annual fatalities per registered vehicle. He noted, however, that other economic variables such as real total personal consumption expenditure and unemployment provided a marginally worse fit for fatality rate.

Joksch (1991) improved Walsh's model by lagging the alcohol consumption variable by two years and retaining a dummy variable for the fuel crises in 1979. The author found that Walsh's computations were incorrect and the model incomplete, and even after making allowances for these problems, concluded that no causal relationship existed between alcohol consumption and the fatality rate.

#### 6.10 Urban planning

Henning-Hager (1986) used a multiplicative regression model to explain the number of accidents in different residential areas within German cities. The explanatory variables were a combination of local and global variables, mostly concerned with the different road characteristics in each region.

The most significant local explanatory variables were found to be: the length of the roadwork; the number of four-or-more directional junctions; a through traffic indicator—the number of possible through routes per thousand inhabitants (pti); lengths of tangential roads pti; and the number of vehicles parked on public roads pti.

The most significant global variables were found to be: the relationship of overall motorization to urban motorization; and the overall area pti.

Fridstrøm & Ingebrigtsen (1991) investigated several aspects of road networks in their compound Poisson-Gamma model of monthly accidents in Norwegian counties. (See section 8.3.) They found that as congestion—as measured by ratio of length of road network in km to gasoline sales from gas stations—

L

increased, the number of crashes decreased. Fatalities decreased more than injuries, and non-occupant injuries more than occupant injuires, which could be due to the reduction in speed caused by congestion.

Broughton (1988) regressed the logarithm of the number of annual fatalities per total traffic volume in Great Britain on a constant trend, the year, and a dummy variable indicating the onset of seat belt use. The value of  $R^2$  for this model was very high (99.5%). The coefficient for year was small, negative and highly significant, indicating a long term gradual decline in the number of fatalities weighted by traffic volume. The author noted that the model only allowed for a linear increase in fatality rate over time.

In a later paper, Broughton (1991), added a quadratic term in year, and an interaction term between year and whether the year was after 1983 or not.

Oppe (1991) studied traffic safety in the Netherlands, in relation to the entire transport system. The traffic system was viewed as a production system, with  $V_t$ , the VKT being the production units, and  $R_t = F_t/V_t$ , the fatality rate per VKT providing an estimate of the probability of failure per unit of production. The output of the system was thus the total VKT, and the total loss on safety was  $F_t$ .

Oppe's model was based on the negative learning model, which supported the theory of social adaptation to traffic put forward by Oppe (1989). (Minter (1987) also investigated learning theory models.) Oppe found that the number of fatalities was a function of the derivative of VKT with a shift in time, and that 95% of the variability in fatlaities could be 'explained' by VKT. The model was used to predict future accident rates.

#### 6.11 Weather as an explanatory variable

#### 6.11.1 Local effects studies

A number of researchers have investigated the relationship between the weather conditions for *specific accidents*, as recorded in accident reports of many countries.

Foldvary & Ashton (1962) mentioned that, for their purposes, the ideal weather information would be the maximum and minimum temperature readings for the period in question, humidity, wind, pressure, visibility (foginess), cloudiness, weather type, existence of other weather phenomena, and rain. Unfortunately, such specific information may only be available for isolated locations within a region, such as large cities. So it is difficult to generalise readings for specific locations to regions.

Johnson & McQuigg (1974) used a principal component technique to model the contribution of rainfall and temperature to average county land prices<sup>3</sup> in the US. Also considered was a linear regression model, with a logit link function, using various explanatory variables to describe the number of fatalities. The climatic explanatory variables used were 4- and 7-day precipitation averages, temperature, log temperature, various combinations of these, or quadratic functions of these variables.

They cited a paper by Benson & Johnson (1970) which considered the problem of measuring economic relationships which include climatic variables, using the method of principal components.

In general, the findings confirmed some early findings by Tanner (1952a, 1952b, 1967).

Wet weather decreased traffic flow but increased the number of accidents and casualties, with the resultant effect of increasing the accident and casualty rates per unit of travel (veh.km). All kinds of traffic were affected, with the greatest reductions in flow in the case of two-wheeled vehicles.

Snow and ice also reduced traffic appreciably, the greatest reductions again occurring in the numbers of two-wheeled vehicles. For accidents, however, the effect depended on the extent of ice and snow: moderate proportions led to more accidents, while larger amounts led to fewer accidents than expected under dry conditions.

Fog reduced traffic appreciably, with much greater reductions at weekends. Accidents overall increased in number, but one class of injury, namely pedestrian, was reduced in number.

<sup>3</sup>This is relevant because of the way in which generalised climatic variables are used to model something.

#### 6.11.2 Global effects models

Foldvary & Ashton (1962) found that the mean sunset time, the number of rainy days, the number of holidays, and a long term trend adequately explained some fortnightly fatality rates from 1960.

Fridstrøm & Ingebrigtsen (1991) were surprised to find that snowfall had a negative effect on monthly crashes in Norwegian counties, and suggested that these results could be explained by any of the following:

- People drive more carefully under adverse conditions.
- Roadside snow drifts would cushion the impact from single vehicle crashes.
- Less people drive in adverse conditions, reducing the exposure to risk of those people who do drive.

All environmental factors could have a worse effect on crashes when unusual or unexpected. The authors measured this 'surprise effect' with a dummy variable indicating whether there had been a snowfall in one month but not in the preceding one. This effect seemed to be offset by the negative effect of total snowfall.

Rainfall was associated with higher crash rates, suggesting that drivers did not appreciate the increasing risk (cf snowfall.)

In addition, the authors found that as the number of daylight hours during rush hours increased, the number of accidents decreased. Note that in Norway, the number of daylight hours may range from 0 to 24 hours during the year.

#### 6.11.3 Generalised climatic variables

There are drawbacks in the structure of information available from a Bureau of Meteorology. The Bureau does not provide a means to generalise the weather conditions for many points within a region, such as a state. Nor do they provide a method for generalising the weather conditions for a particular point over a period of time. The latter problem may be summarised by two types of statistics: the measure of central tendency (mean, median or mode); and a measure of variability (range, upper and lower quantiles, or histogram.) Maunder (1974) considered the problem in depth.

He formulated periodic (say monthly) weighted indices for climatic variables such as rainfall and temperature. The weightings were based on the contribution of a region to the national total population or area. Other variables suitable for our application to road safety would be the number of registered vehicles, vehicle kilometres travelled, and the kilometres of road in a particular region, such as a state.

In this study, climatic variables, such as rainfall, were measured at a number of stations within a region, where long-term (say twenty year) average values were known for each station. The rainfall for station i, expressed as a percentage of the long-term average for that station, was denoted  $C_i$ , and the percentage of the regional road safety parameter (such as population, land area, road length) in area i, was denoted  $E_i$ . The regional Climatic Index, I, was defined as:

$$I = \frac{\sum C_i E_i}{\sum E_i}$$

In answer to skeptical meteorologists, he stated that 'it could of course be argued that a weather index for a nation as large as the US has little physical or practical meaning. Nevertheless, it is strongly believed that if some measure of nation-wide weather can be computed, and that it can be of use to decisionmakers.'

#### 6.11.4 More recent applications of generalised indices

Scott (1986) used weather variables in regression and Box-Jenkins models of monthly accident data in Great Britain. The author stated that 'monthly data are available, summarising temperature and rainfall throughout Great Britain.' Unfortunately, we are given no indication of how the summaries were obtained.

In the regression analysis, including several variables, the authors found that high rainfall and warmer temperature were related to high accident frequencies, as was expected. Similar results were obtained when an ARIMA error term was used in the model.

# 6.12 Fuel prices

Several authors have studied how fuel prices have affected road crashes. Allsop & Turner (1986) used the real level of pump prices for petrol to model monthly fatalities from 1978 to 1983 in Great Britain, disaggregated by road user type. MacLean (1983) found that a fast increase in fuel prices had a short term 'shock effect' as well as a long term effect which could not have been due to gradual increase in prices, and suggested that catastrophe theory should be investigated. Scott (1983) found a strong relationship between crashes and petrol prices. ARIMA and structural time series models were used by Harvey & Durbin (1986) to model fatal crashes. They used just two explanatory variables: a car traffic index, and the real price of petrol in their final explanatory model.

#### 6.12.1 Predicting Fuel Prices

Fuel prices were often used in models used to predict future levels of traffic safety, so the problem of predicting fuel prices became important.

In the paper of Donaldson, Gillan & Jones (1990), future annual consumption rates in Australia were estimated as a function of fleet size and average vehicle fuel consumption. Fleet size is predicted using a regression of population on time. Average vehicle fuel consumption is determined 'econometrically' from fuel price.

Wheaton (1982) used component estimation to predict future gasoline consumption from variables such as average fuel efficiency, fleet size and VMT. However, Cervero (1985) pointed out that although econometric models such as the above model major structural features and turning-points well, short-term estimation is difficult since the explanatory variables are themselves difficult to predict.

Thus, in order to forecast monthly highway energy consumption in the US, Cervero (1985) used ARIMA models to avoid including more explanatory variables. Monthly and biannual seasonal factors were useful in short-term forecasting. The oil crises of 1979 were considered but found to have no significant impact on results. Disaggregated data was also considered.

- --

# 7 Time series analysis

The number of fatal crashes or fatalities measured over time may be correlated, as many variables measured over time often are. ARIMA models take into account this autocorrelation of the series, and are specially designed to model seasonal and long-term trends. The classic volume by Box & Jenkins (1976) introduced and popularised this particular approach to modelling time series. Bhattacharrya et al (1979), an Australian study, was one of the first road safety studies to utilize the Box-Jenkins models.

To analyse the evolution of a road safety indicator over time, the more common method employed in the road safety literature is a regression analysis, using dummy variables to indicate a trend over time, and perhaps a seasonality factor. See the introductory section on Regression Analyses, §4, for some examples. However, note that regression models assume independence between error terms.

# 7.1 Comparison of uncorrelated normal and ARIMA error structures

Lassarre (1986) conducted a study of the monthly number of accidents and number of fatalities on roads covered by the Gendarmerie Nationale in France (mostly nonurban roads) between 1970 and 1977, as related to the introduction of speed limits and compulsory seatbelt wearing. A major problem with data collection was obtaining external variables measured during the same time period, with the same periodicity, and for the same location. The explanatory variables considered were: the monthly traffic volume index, an annual series of speeds for light vehicles and an annual seatbelt wearing index. Annual series were converted to monthly series by fitting curves and interpolating.

#### 7.1.1 The regression model

The model used was a log-linear regression with an ARIMA error term. Assume that all variables, apart from the error structure are logged in the following model equation.

$$D_{t} = v_{0}V_{t} + \alpha_{1}\bar{S}_{t} + \alpha_{2}se(S_{t}) + \alpha_{3}B_{t} + (w_{0}w_{1}B)\xi_{42,t} + w_{2}\xi_{48,t} + \frac{(1-\theta_{1}B)(1-\theta_{12}B^{12})}{(1-B)(1-B^{12})}a_{t}$$

where

Ł

$D_t$	=	number of deaths in month t
$V_t$	=	traffic volume index
Ŝ,	=	average speed in month t
$se(S_t)$	=	standard error of speed in month t
$B_t$	=	ratio of seatbelt wearing
2nd part	=	sum of two dummy variables indicating beginning of safety measures in June, July and Dec 1973
3rd part	=	autoregressive structure of remaining error

Another similar study, Scott (1986) examined a monthly accident series in Great Britain. A basic regression model with log terms was fitted first:

$$\ln A = v_1 \ln V_1 + v_2 \ln V_2 + p \ln P + tT + dD + rR + wW + f\delta F + s\delta S + k + e$$

where		
A	=	the number of accidents
$V_{1}, V_{2}$	=	traffic volume indicators
Р	=	the petrol price index, the ratio of average monthly retail price to the overall retail price index
T	=	offset in months from the beginning of the time period
D	=	temperature (deg C)
R	=	rainfall (mm)
W	=	the number of working days in a month
δF	=	a dummy representing prescence of the fuel crisis (Dec 73-Apr 74)
85	æ	a dummy representing prescence of legislation for lower speed limits outside built-up areas (Dec 74-May 77)
k	=	a seasonal factor and
е	=	an error term

#### 7.1.2 Regression results

The main results are described in more detail in Scott (1983):

- There was strong evidence against a simple linear relationship existing between accident frequencies and traffic volumes.
- Petrol price was highly related to accident frequencies, except for twowheeled vehicles (as to be expected.)
- High rainfall and warm temperatures were associated with higher accident frequencies.
- The series of most explanatory variables exhibited trends, generally downward and steady over the period.

An analysis of residuals indicated that a good fit had been obtained. A few series had autocorrelated residuals, suggesting that time-series models might perform better.

#### 7.1.3 Model with ARIMA error term

Secondly, an ARIMA model of the error term was considered. All of the variables remained in the new model, except the time and seasonal factors, which were incorporated into the error structure:

 $\ln A = v \ln V + v_2 \ln V_2 + p \ln P + dD + rR + wW + f\delta F + s\delta S + e$ 

where  $e = ARIMA(\epsilon)$ , and  $\epsilon$  is white noise or normal error.

#### 7.1.4 Comparison

Similar estimates for coefficients were obtained for both types of model. Although the Box-Jenkins model performed slightly better, the simplicity of the ordinary regression model overcomes this slight disadvantage.

#### 7.2 Intervention analysis

Votey (1986) discussed the advantages and disadvantages for using various models for road accident frequencies. Intervention, ARIMA and simultaneous regression techniques are described in broad detail, with a particular model being favoured for its ability to relate to the underlying theories and characteristics of road safety. The models were:

$$DD = d(ALC, PA, SV, KD, ...) \qquad PA = p(DD, RQ, ...)$$

where DD is the level of drunken driving, ALC is alcohol consumption, PA is the probability of apprehension and sanctioning, SV is the severity of sentence, KD is the distance driven, RQ is road quality, for some structural models p and d. Undoing the recursion yields the following relation:

$$ALC = a(DD, KD, VM, RQ, ...)$$

where VM is vehicle mix.

# 7.3 Identification of unknown intervention times in time series

The paper by Helfenstein (1990) is a continuation of other papers which use traditional time series analysis methods, such as Box-Jenkins and structural time series models. These preliminary papers are: Bhattacharyya et al (1979), Box & Tiao (1975), Harvey & Durbin (1986), and Lassarre & Tan (1982).

Helfenstein's paper was slightly different in that the intervention time was assumed unknown. In the statistical literature, this is known as the changepoint problem. The change-point may be deduced from the data by using a number of both graphical and numerical techniques. If the actual time of an intervention is close to that identified by the above techniques, then this may provide more evidence for the efficiency of the countermeasure.

The example data used were quarterly numbers of accidents with injury in Zurich from 1976 to 1985.

The different techniques are outlined below.

Plot of seasonal subseries This is constructed by joining together observations from each season. Thus, four curves will be produced for quarterly data, twelve for monthly data, etc. The first and last dates that an decrease or increase occurred will be indicated where each curve begins to 'drop or bounce'.

Series of seasonal differences For annual or nonseasonal data, the series of first differences,  $\nabla y_t = y_t - y_{t-1}$  should contain a spike at the time of intervention. For seasonal data

$$y_t = \beta \xi_t + s_t + \epsilon_t$$

where  $\xi_t = I[t \ge T]$ , T is the 'unknown' moment of intervention,  $\beta$  is the height of the step,  $s_t$  is a periodic function with seasonal fluctuations ( $s_t = s_{t-4}$ ) for quarterly data, and  $\epsilon_t$  is 'unexplained' variation.

Or alternatively,

$$\nabla_4 y_t = \beta \nabla_4 \xi_t + \nabla_4 \epsilon_t$$

where  $\nabla_4 \xi = I[T \le t \le T+3]$ 

**Cross-correlation function** The more statistical methods are based on first fitting a simple ARIMA model (also called Box-Jenkins models) to account for the long-term trend and any seasonal effects. For example

$$\nabla_4 y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^4)a_t$$

takes account of a quarterly seasonal effect, and an autocorrelation of values.

After fitting the ARIMA model, Helfenstein suggested the calculation of the cross-correlation function between the fitted values (reference signal) and the series of residuals. If the fitted values were:

$$p_t = \begin{cases} 0 & t < \text{intervention time} \\ -1 & t >= \text{intervention time} \end{cases}$$

then the cross-correlation function  $r_{pa}(k)$  between the fitted values and the residuals  $a_t$  would be:

$$r_{pa}(k) = \operatorname{correlation}(p_t, a_{t+k})$$

The cross-correlation function should be calculated for all possible times of intervention. The maximum correlation between  $p_t$  and  $a_t$  will give an estimate of the intervention.

Residual variances of successive interventional models The intervention analysis presented by Box & Tiao (1975) is of the form  $y_t = \beta \xi_t + z_t$ , where  $\xi_t = I[t \ge T]$ ,  $\beta$  is the unknown impact of the intervention, and  $z_t$  is an ARIMA process.

For each possible intervention time T, an intervention analysis may be performed and the resulting residual variances plotted against the intervention time.

If, for example, the number of accidents decreases before the actual intervention time, then an anticipatory (pre-intervention) effect may be fitted in addition to an intervention effect, involving one more  $\beta_2 \xi_{2t}$  term in the model, corresponding to the second important timepoint.

#### 7.4 ARIMA model, with explanatory variables

Wagenaar (1984) used an ARIMA model, unemployment and VMT to model monthly numbers of crash involvement in the state of Michigan.

First, negative correlations between unemployment and crash involvement were found to be significant, a lagged relationship was also found to be significant. Then the author fitted the unemployment series with an ARIMA model, incorporating first and seasonal differences. This model was then applied to the crash involvement data, and the residuals from the unemployment and crash involvement series cross-correlated. The first and second lag of unemployment and the seasonal and first difference component of crash involvement were retained in the resulting model.

The same procedure was followed to analyse the relationship between VMT and CI (crash involvement.) Although a strong relationship was expected between the two variables, the extent of the lag required for this to be evidenced was not expected by the author. The final model included an AR(2), AR(3) term for VMT, and a seasonal component and first difference for CI.

Since unemployment and VMT were independent of each other, as indicated by cross-correlation calculations, the next step was to include them both in the model. The final model included first differences and seasonal differences for CI, trend and first lag of unemployment, and first and second lag of VMT.

#### 7.5 Structural time series models

Harvey & Durbin (1986) considered the use of structural error terms. In a structural time-series model, the trend, moving average and autoregressive portions of the model may all involve a stochastic random walk component. Thus the parameters of the model could vary with time, instead of being fixed, as is

generally the case. Time-varying parameters enhance the explanatory power of the model, although they decrease the predictive power of the model, due to the added 'randomness' introduced. This is discussed in more detail in Volume 2 of this report.

The paper's primary concern was the effect of seat belt legislation on the numbers killed and seriously injured (KSI) each month in the UK for different types of road users. The authors compared structural time series models to ARIMA models.

The form of their final model was

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^k \delta_j \boldsymbol{x}_{jt} + \lambda \boldsymbol{\omega}_t + \boldsymbol{\epsilon}_t$$

where the level and slope of the general trend  $\mu_t$  were determined by random walks:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \qquad \beta_t = \beta_{t-1} + \zeta_t$$

The seasonality was modelled by

$$\gamma_t = \sum_{j=1}^{j/2} \gamma_{jt}$$

where for s even and  $\lambda_j = \frac{2\pi_j}{4}, \quad j = 1, \dots, \frac{j}{2} - 1$ ,

$$\begin{bmatrix} \gamma_{jt} \\ \gamma_{jt}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix} \begin{bmatrix} \lambda_{j,t-1} \\ \lambda_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega_{jt}^* \end{bmatrix}$$

where

$$\gamma_{jt} = (\cos \lambda_j) \gamma_{j,t-1} + \omega_{jt}, \qquad j = \frac{s}{2}$$

and the  $\{w_{jt}\}$  and  $\{w_{jt}^*\}$  were *ind* N(0,  $\sigma_{\omega}^2$ ).

It was also possible to allow  $\{\omega_{jt}\}$  and  $\{\omega_{jt}^*\}$  to vary with j, which permitted the seasonal pattern to vary with time. The value of the *j*th explanatory variable at time t was  $x_{jt}$  and  $\delta_j$  was its coefficient (not time-dependent). The intervention variable,  $\omega_t = 1$ ,  $t \geq \tau$ ; 0 otherwise.

The evaluation statistics used by the authors were:  $\bar{\sigma}^2$ , the estimated one step prediction error variance; the usual  $R^2$ ;  $R_i^2$ , which is  $R^2$  adjusted for seasonality; H, the heterogeneity test statistic; Q(P), the Box-Ljung statistic on the first P autocorrelations of the standardised residuals; a Normality test statistic for the residuals; a Post-sample predictive test statistic used to test goodnessof-fit; CUSUM, a test which detects model breakdown; and the Recursive t-test, which may be used if it is expected that the residuals after the intervention may all be of the same sign.

The authors decided to use a log transformation since they were using count data; they also considered the square root transformation. They considered just two explanatory variables: fuel prices and a traffic index. Finally, they found that for road users who would be directly affected by seatbelt legislation, the number of deaths decreased: there was an estimated 18% reduction in deaths for drivers, and a 25% reduction in deaths for front seat passengers. For those not directly affected by the legislation, there was a highly significant 27% increase in deaths for rear-seat passengers, and not very significant increases in deaths for passengers (8%) and cyclists (13%).

It is interesting to note, however, that the authors found that a simple 'airline' model was adequate for prediction purposes, although not for explanatory purposes. They found that a version of the airline ARIMA model fitted the data well:

$$(1-B)(1-B^{12})y_t = (1-.684B)(1-.995B^{12})\xi_t, \quad \bar{\sigma} = 0.075$$

However, they stated that "The structural approach that we have adopted represents, we believe, a more direct and transparent technique for time series modelling."

Another paper, Wilson (1986), used an airline ARIMA model, and found little difference between it and a 'structural' model. Wilson preferred the structural model since it used explanatory variables to predict the response variable, to support any cause-effect hypotheses, whereas the ARIMA model merely used the past history of the response variable to predict its future values. Regression models with an ARIMA model term were not considered by this author.

Martinez-Schnell & Zaidi (1989) investigated the daily, weekly and monthly time series of deaths due to six different types of injuries: motor vehicles, suicides, homocides, falls, drownings, and residential fires. Motor vehicle deaths, the major class of deaths, were investigated in more detail, using transfer functions and intervention analysis.

They investigated several calendar effects variables:  $d_t$ , the number of days in the month;  $w_t$ , the number of Saturdays and Sundays in the month;  $h_t$ , the number of holidays in the month. Other explanatory variables considered were  $x_t$ , the VMT in the month;  $I_t$ , an indicator variable for the oil crisis,  $I_t = 1$  for the months in 1974–1983, and is 0 otherwise.

An ARIMA model similar to an airline model was fitted to VMT. This model for  $x_t$  was used in a transfer function to model  $y_t$ , the number of deaths due to road crashes in a month.

The best model for  $y_t$  included a transfer function for  $z_t$ , the intervention variable for the oil crisis,  $I_t$ , and two of the calendar effects variables,  $w_t$  and  $h_t$ .

The inclusion of  $x_t$  in the model reduced the residual variation by 32%, and its coefficient was significant (t > 11), implying that for every 100,000 mile increase in VMT per month, there was a corresponding increase of approximately 3 deaths. The variable  $d_t$  was dropped from the model after  $x_t$  was included. This can be explained by correlation between VMT and the number of days in the month—the more days there were in the month, the more miles were travelled on the roads. The intervention variable accounted for a 33% drop in the residual variance after  $x_t$  was added, and was significant (t = -2.23). The model suggested that there was a significant drop of 352 deaths during the oil crisis, even after accounting for VMT and calendar effects.

Seasonal variables were not found to be significant. The authors used the mean square error,  $R^2$ , the Box-Ljung  $\chi^2$  statistic, and the Aikeke Information Criterion to evaluate the goodness-of-fit of the various models. They also constructed weekly and daily time series of the number of deaths. Their week started on Wednesday and ended on Tuesday so that weekends and Fridays and Mondays, which were often public holidays, were not split up. The weekly and daily time series from the monthly time series.

# 8 Other statistical models

So far, various regression and time series models have been suggested for the road crash data. Some alternative models that have been presented are described below.

#### 8.1 Learning-theory models

Road safety may be viewed as a 'learning process' which society is undergoing according to Minter (1987). Two different models for learning processes were discussed in this paper, where in learning theory, a certain 'task' is to be learned by society. This task is repeated a number of times, and becomes easier as experience increases.

Wright's model is

$$t_n = t n^{-b}$$

where  $t_n$  is the time for the *n*th task performed; *n* is the cumulative number of repetitions so far, the measure of experience; *b* is a measure of the rate of reduction of time per repetition, usually  $0.7 \le b \le 1$ .

Torwill's model is

$$y_n = y_0(1 - \exp\left(1 - \frac{n}{t}\right)) + c$$

where  $y_n$  is the measure of performance at time *n* periods after the start;  $y_0$  and *c* are constants; and *t* is time.

When these models were applied to road safety data, the measure of performance was the casualty rate, expressed as casualties per distance travelled for Wright's model, and its inverse in Torwill's model. Torwill's time parameter and Wright's number of repetitions were the measures of experience. Minter suggested that vehicles per head of population was a good measure of accumulated experience. Note that Wright's model is analogous to Smeed's formula, where the measure of performance is deaths per vehicle.

Finally, Minter found that both learning curves fit the data well, concluding 'that things will improve anyway, and that left to themselves they might even improve better than with legislation intervention.'

Oppe (1991) also used a model based on learning-theory to explain road fatalities in the Netherlands. See section 6.10.

#### 8.2 A systems-based model

Blomquist (1986) employed 'consumer utility theory' to model motor vehicle accidents. The thoery is described in more detail in the paper referenced, as well as Hakim *et al* (1991) but is described briefly below.

The theory is based on the application of the benefit-cost approach to the individual driver, who may trade off between reduced risk (more safety) and

increased utility (more cost or effort). However, since not all drivers aim to reduce the risk to every driver, a system of penalties, such as Pigorian prescription, may be used. This system aims to alters the individual utility function to account for the safety of society in general.

However, utility is difficult to measure, so an indirect measure may be used, such as accessibility (or mobility) which affects utility. For example, increasing the price of alcohol, would reduce its availability, or increase the penalties for DUI, and thus make it more difficult for the driver to drink and drive, which would therefore increase the utility of DUI.

Accessibility is affected by both society and the individual. Society may affect accessibility by changing the transport infrastructure, and level of police enforcement. Individuals' behaviour may affect the level of gasoline consumption, investment in vehicles, number of drivers on the road, and the time spent driving. Alternatively, Hakim et al (1991) note that a demand model for accessibility might be based on income (GNP), travel price (price of gasoline) and relative prices of alternative modes (price of public transport.) See Allsop and Turner (1986) also for a discussion of this.

Finally, Ilakim et al (1991) conclude that the aim of society is to maximise accessibility whilst maximising safety, by minimising the number of accidents. Accidents in turn may be explained by levels of exposure (VMT, gasoline consumption), social norms and behaviour (rates of crime, suicide), legislation (speed limits, vehicle inspection, seatbelt laws, police traffic enforcement, punishment for offences.) Hence, in order to work out how to minimise accidents, the whole system surrounding the production of accidents and the environment in which they occur must be modelled.

# 8.3 Poisson distribution

Although it is fairly common to model the number of accidents per time period as a Poisson process (see Weed (1986), Jadaan & Salter (1982) for example), the number of fatalities or casualties is much harder to model. MacLean & Teale (1982) partially investigated a compound model based on the Poisson distribution of the number of accidents. They defined X to be the number of accidents and assume that it was distributed as  $Poisson(\lambda)$ . The conclusions deduced by the authors were that the possible values for Y comprised the nonnegative integers, and that the variance of Y was larger than that for a Poisson distribution.

Fridstrøm & Ingebrigtsen (1991) applied a compound Poisson-Gamma model, a special case of the Generalised Linear Models of McCullagh & Nelder (1983) to several cross-sectional and longitudinal measures of traffic safety. The number of crashes was modelled as a negative binomial distribution with an expected value whose logarithm was a linear function of the explanatory variables and a Gamma random variable. Thus, the variance was larger than the expected value, which seems to be the property required by MacLean & Teale (1982). The form of the explanatory variables was first suggested by Gourieroux et al (1984).

# 8.4 Poisson regression

Jovanis & Chang (1986) review some techniques used to model the number of accidents according to a Poisson distribution and found that:

- If accident frequency is regressed on VMT, and if accident frequency really is Poisson, then the variance of accident frequency will increase as VMT increases.
- If a square root transformation is performed on the accident frequencies, a common antidote for Poisson data, then the variance problem will be addressed, but this is accompanied by a bias in the estimates.
- Because accident frequencies are nonnegative by nature, this can create problems with unconstrained least squares analysis. Constrained least squares can overcome this problem, but also results in biased estimates. Transformation of the accident frequencies to a non-linear form may solve the nonnegativity, but generally introduces discontinuities; for example, log 0 is undefined.

#### 8.5 Other distributions

Weed (1986) referred to the method described in the Highway Safety Evaluation Procedural Guide for comparing the accident count before and after some countermeasure has been applied. The author used a Poisson, ChiSquare, Binomial and a modified Binomial statistics based on these before and after accident counts.

The number of accidents which occur after the countermeasure has been applied,  $O_1$ , and the number of accidents occurring before the countermeasure,  $O_2$ , can be compared using the  $\chi^2$  goodness-of-fit statistic. The authors set the expected values of  $O_1$  and  $O_2$  to be their average,  $\frac{(O_1+O_2)}{2}$ .

Alternatively, the number of accidents before the countermeasure is applied may be modelled as a binomial distribution, with parameter N, the total number of accidents sampled being the number of accidents before and after the countermeasure. Under the null hyothesis that the number of accidents before and after are similarly distributed, the parameter p is 0.5. A modified binomial method ensures that, with discrete data, the more conservative estimators are always used.

Weed performed a thousand simulations of one Poisson process to show that the standard process suggested by the HSE procedural guide, using regression, produced highly optimistic estimates of whether the before and after counts were equal. The modified binomial was the most accurate and both the  $\chi^2$  and the binomial methods produced slightly conservative estimates.

The number of accidents before and after some intervention was also modelled with Poisson, ChiSquare, Binomial and modified Binomial statistics by Jadaan & Salter (1982). Jovanis & Chang (1986) modelled the number of accidents by a Normal distribution, and noted that the usual regression by least squares assumes Normal errors as justification for this. They also considered the use of Bernoulli trials (success if a person completed a trip without injury) and survival theory as an extension of this.

# 8.6 Epidemiological approach

One way to look at road crash fatalities is to call them a 'disease', like any other which kills a lot of people. Researchers have utilized methods, or ideas, from conventional epidemiological research, and applied them in this context. An overall view of Australian mortality is provided in Spencer (1980).

Many national government bodies are interested in the mortality rates due to different causes, including road crashes. Comparisons of death rates in various countries attempt to place a particular country's safety rating within context.

#### 8.6.1 International aggregates

Periodically, countries overseas publish these aggregates, stratified by several indicators such as road user type, age and sex of driver, etc. Japan, America, Australia, African countries, and the European countries are among those countries who publish these yearly statistics. World Statistics quoted only the most extreme values for percentage increase in the number of accidents, casualties and fatalities. Hutchinson (1987) is a very thorough compilation of international road accident statistics, which also described both the official and alternative sources, whence these statistics may be obtained.

#### 8.6.2 Comparing international statistics

Another use for international data is to provide a 'benchmark' to compare the safety standards in Australia to various other countries. Several studies have already attempted to compare the fatality rates of different countries. For example, a study by Berard-Andersen (1978) compared the use and effects of seat-belts in twenty-one countries.

A method which has been used to compare traffic safety between countries was first introduced by Smeed (1949). This will be discussed in more detail in the following section.

#### 8.6.3 Risk and exposure

The use of hazard models, and the ideas of risk and exposure are relatively well-developed within the field of epidemiology.

# 8.7 Discriminant analysis

Neuman (1984) applied discriminant analysis to pick out variables which contributed to high accident sites, as compared to low accident sites. This technique could conceivably be applied to differences in fatality rates between states, or regions within states.

Ł

# 9 List of explanatory variables

This list indicates which variables have been considered in models for fatalities (or fatality rates) by various authors. It is noted whether the variables contributed significantly to the final model proposed by the author. However, it is necessary to inspect the variables in context, within the framework of the model together with companion explanatory variables, to fully understand their significance.

Variable	Ref	Response	Sig effect?
r ·			
Legislation sent helts	[4.9]	local acc in Via 1963	nogotive
seat Dens	[44] [199]	fot noto per registered	negative
dummy var.	[122]	veh.	no
change in	[31, 39]	fat. per veh. 1984, coun-	negative
urban/rural speed		try	
limits.	[no]		, <del>.</del>
change in posted speeds	[22]	lat. rate per population	negative
speed limit dummy	[38]	annual fat. for	negative
variable		cars/trucks	
maximum speed	[70]	1970 state fat. per VMT	varying
limit		per popn, per veh.	
ratio inpected vehi-	[38]	annual fat. for	no
cles		cars/trucks	
inspection dummy	[80]	annual fat. New Jersey	yes
variable	[70]	annual fat	in conclusive
venicle inspection	[19]	annuar lat.	meonensive
introduction of	[122]	iat, rate per registered	no
dummy var		ven.	
nolice activity	[130]	annual	ло
minimum driving	[70]	1970 state fat ner VMT	verving
age	ניטן	per nonn. per veh.	varynig
driving age	[73-8]	au <b>nua</b> l	negative
drinking age	[8]	annual	conflicting
drinking age	្រៀ [ទក]	onnual	conflicting
urinking age	[00] (100]	annual	indinest per
liquor taxes	[122]	annual lat.	nullect neg-
			GUIAC

ι

Variable	Ref	Response	Sig effect?
Economic factors			
fares	[4]	fatal-serious accidents per month	
pump petrol prices	[4]	fatal-serious accidents per month	
fuel prices	[83]	fatalities	
recession	[29]	fatal accidents per reg. veh.	negative
cost of an accident	[38]	annual fat. for cars/trucks	negative
accident cost	[102.37]	annual	negative
real dis- posable income per driver (\$1972)	[38]	annual fat. for cars/trucks	no
income	[68]	annual fat.	not best
real PC annual in- come	[80]	annual fat. New Jersey	yes
real average weakly earnings	[117]	annual Aust.	some
Gini index of in- come	[130]	motor veh. related mor- tality, different countries	some
income	[102,134,	,130]	
		annual .	positive in longitudinal
income	[102,134	,130]	
		annual	negative in crosssec- tional
Federal Reserve Board Index of In- dustrial Production	[88]	annual fat.	best
annual % change in real state GDP	[117]	annual Aust.	some
PCGNP	[130]	motor veh. related mor- tality, different countries	some
per capita GNP	[130]	annual	positive up to a point, and then negative

Variable	Ref	Response	Sig effect?
Economic factors, con	ntinued		
unemployment rate	[68]	annual fat.	not best
unemployment	[101,120]	annual	positive
automobile produc-	[68]	an <b>nual fat</b> .	not best
tion rate			
motorization rate	[90]	fat. for one year for dif-	yes
	[ ]	ferent countries	
WWII dummy vari-	[80]	annual fat. New Jersey	yes
able	[76]	annual	few rds
nospital access	[, ]		

ļ

Variable	Ref	Response	Sig effect?
Driver demographics			
strike days	[4]	fatal-serious accidents per month	
change in industrial activity	[68]	annual	negative
holiday	[32]	local fortnightly acc 1960	yes
alcohol	[38]	annual fat. for	no
consumption		cars/truc <b>ks</b>	
PC alco-	[85]	annual fat. Ontario	yes
hol consumption for persons over 14 yrs			
beer consumption	[121]	serious injuries	yes, also lag 1
alcohol	[122]	fat. rate per registered	yes
consumption PC for		veh.	
persons over 14 yrs			
alcohol	102,103	8,67,25,132,134,37,80]	
consumption			
		annual	positive
ratio youths	[38]	annual fat. for cars/trucks	positive
%young drivers	[121]	moderate injuries	yes
youth	102,67	25,8,38]	
•		annual fat. rate	positive
male	73-	annual	positive
%male drivers	Û,	1970 state fat per VMT.	•
		per popn, per veh. vary- ing	
seatbelt wearing	74	monthly fat. 1970-77 France	yes
drivers DUI	[85]	annual acc Ontario	yes
% and	<i>[91]</i>	monthly fat. Canada	no
DUI charges per month	[91]	monthly fat. Canada	по
citations per driver	[76]	annual, by roadtype	yes most roads

Variable	Ref	Response	Sig effect?
Vehicle demographics			
urbanisation	[33]	fat. per person Aust	positive
urbanisation	[70]	1970 state fat. per VMT, per popn, per veh.	varying
ratio auto- mobiles with safety equipment	[38]	annual fat. for cars/trucks	negative
automobile safety regulation	[102,25,4	1,132,134,37,38]	
		negative	
traffic volume index	[74]	monthly fat. 1970-77 France	yes
mean speed	[74]	monthly fat. 1970-77	yes
mean speed	[76]	annual	not some roads
average speed	[102.67.1]	07.8.80.132.134.37.38]	10803
arouge spece	[,	annual	not really
se(speed)	[74]	monthly fat. 1970-77 France	yes
se(speed)	[76]	annual	yes
speed variability	[76]	annual	positive
vehicle mix	[130]	annual	
motor veh. registra- tions	[80]	annual fat. New jersey	yes
acc. involving sin- gle vehicles	[85]	annual acc Ontario	yes
veh. stopped by BATmobiles	[91]	monthly fat. Canada	no
highway capacity	[70]	1970 state fat. per VMT, per popn, per veh. vary- ing	

Variable	Ref	Response	Sig effect?
Ezposure			
number of vehicles	[2]		
population	[2,80]		yes
annual $\%\Delta$ VMT	[68]	annual fat.	yes
miles highway	[38]	annual fat. for cars/trucks	negative
VKT	[117]	annuai	
VMT	[121]	property-related crashes	yes
car registrations	[51]	annual accident rate	yes
length road network	[55]	annual accident per pop- ulation	yes
possible routes per population	[35]	annual accident per pop- ulation	yes
area in hectares	[55]	annual accident per pop- ulation	yes
drunk-driving	[91]	monthly fat. Canada	no
related news-			
paper/magazine ar- ticles			

•

Variable	Ref	Response	Sig effect?
Weather			
temperature	[65]		
annual average teu-	[70]	state 1970 fat. ner VMT.	
perature	[, •]	per popn, per veh. vary-	
<b>F</b>		ing	
temperature	[73, 134]	annual	positive
rainfall	[87]		
rain	[22]	local effects	yes
rain	[32]	local hourly acc in 1960	positive
rainy days	[32]	local fortnightly acc 1960	yes
days snow	[4]	fatal-serious accidents	
		per month	
snow	[22]	local effects	yes
fog	[22]	local effects	no
wet road conditions	[22]	local effects	yes
wet weather index,	[57]	local effects on fat.	yes
based on skid resis-			
tance			
icy road conditions	[22]	local effects	no
nighttime fat. acc.	[85]	annual acc. Ontario	yes

# References

- J. Adams. Smeed's law, seat belts and the emperor's new clothes. Human Behaviour and Traffic Safety, 1985.
- [2] J. Adams. Smeed's law: some further thoughts. Traffic Engineering and Control, 28(2):70-74, 1987.
- [3] B. Aldman. Traffic accidents. Transporthaelsan, 57-61, May 1980.
- [4] R. E. Allsop and E. D. Turner. Road casualties and public transport fares in London. Accident Analysis and Prevention, 18(2):147-156, 1986.
- [5] D. C. Andreassen. Linking deaths with vehicles and population. Traffic Engineering and Control, 26(11):547-9, November 1985.
- [6] D. C. Andreassen. Population and registered vehicle data vs. road deaths. Accident Analysis and Prevention, 23(5):343-351, 1991.
- [7] R. Arnold and E. C. Cerrelli. Holiday effect on traffic fatalities. Technical Report DOT-HS-807115, NHTSA, April 1987.
- [8] P. Asch and D. T. Levy. Does the minimum drinking age affect traffic fatalities? Journal of Policy Analysis and Management, 6:180-192, 1987.
- [9] F. J. Benson and S. R. Johnson. Principal components and problems of measuring economic relationships which include climatic variables. In Proceedings, 2nd National Conference on Weather Modification, American Meteorological Society, pages 419-26, 1970.
- [10] K. Berard-Andersen. Use and effects of seat-belts in 21 countries. Monograph HS-033 468, Institute Transport Economics Norway, Jan 1978.
- [11] M. N. Bhattacharrya and A. P. Layton. Effectiveness of seat belt legislation on the Queensland road toll. An Australian case study in intervention analysis. Journal American Statistical Association, 74:596-603, 1979.
- [12] G. C. Blomquist. A utility maximisation model of driver traffic safety behaviour. Accident Analysis and Prevention, 18:371-375, 1986.
- [13] G. E. P. Box and G. M. Jenkins. Time series analysis: forecasting and control. San Francisco, C-A: Holden-Day, 1976.
- [14] G. E. P. Box and G. C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal American Statistical* Association, 70:70-79, 1975.
- [15] C. P. Brinkman. Toward improved highway safety analysis. Public Roads, 50(2):49-54, September 1986.

- [16] R. G. Buxbaum and T. Colton. Relationship of motor vehicle inspection to accident mortality. The Journal of the American Medical Association, 197:101-6, 1966.
- [17] A. Ceder and M. Livneh. Relationships between traffic flow and hourly traffic flow—I. Accident Analysis and Prevention, 14:19-34, 1982.
- [18] E. C. Cerrelli. Driver exposure: the indirect approach for obtaining relative measures. Accident Analysis and Prevention, 5:147-156, 1973.
- [19] Robert Cervero. Short-run forecasting of highway gasoline consumption in the United States. Transportation Research - A, 19A(4):305-313, 1985.
- [20] H. Chang. A disaggregate survival model of motor carrier highway accident occurrence. Northwestern University, Department of Civil Engineering, 1987.
- [21] Hsin-Li Chang and Paul P. Jovanis. Formulating accident occurrence as a survival process. Accident Analysis and Prevention, 22(5):407-419, 1990.
- [22] P. J. Codling. Weather and road accidents, chapter 11. Newton Abbott, 1974.
- [23] P. J. Cooper. The nature and causes of the 1982 traffic accident reduction in British Columbia Technical Report, Insurance Corporation of British Columbia, March 1984
- [24] D. R. Cox. Regression models and life tables. Journal of the Royal Statistical Society, 2,187-220, 1972.
- [25] Robert W. Crandall and John D. Graham. Automobile safety regulation and offsetting behaviour: some new empirical estimates. AEA Papers and Proceedings, 74(2):328-331, 1984.
- [26] Stig Danielsson, A comparison of two methods for estimating the effect of a countermeasure in the presence of regression effects. Accident Analysis and Prevention, 18(1),13-23, 1986.
- [27] P. Kim Donaldson, Paul Gillan, and Barry P. Jones. Road transport fuel demand in australia: projecting passenger vehicle fuel use. In Conference of Economists, UNSW, Sydney, September 1990.
- [28] M. Ensenberg. The analysis of accidents and casualties for the years 1976-83, with special reference to the Easter and Christmas period. Tech Report RY/8/84, NITRR, October 1984.
- [29] J. M. Eshler. Filtering of fatal-accident rates. Transportation Research Record, 643:10-12, 1977.

- [30] J. C. Fell. Alcohol involvement in traffic accidents: recent estimates for the National Center for Statistics and Analysis, Statistical report. NRD-32 HS-806269, NHTSA, May 1982.
- [31] R. Fieldwick and R. J. Brown. The effect of speed limits on road casualties. Traffic Engineering and Control, 28(12):635-40, December 1987.
- [32] L. Foldvary and H. Ashton. Road accidents and weather. In Proceedings, 1st ARRB conference, pages 529-583, 1962.
- [33] L. Foldvary, C. Watkins, and M. Cameron. Australian road accident rates. Technical Report AIR-016, ARRB, 1971.
- [34] J. C. Fortenberry, D. B. Brown, and D. M. Ryhne. A non-parametric quasi-experimental technique for evaluating highway traffic safety countermeasures. *Transportation Research Record*, 1047:18-20, 1985.
- [35] Lasse Fridstrøm and Siv Ingebrigtsen. An aggregate accident model based on pooled, regional time-series data. Accident Analysis and Prevention, 23(5):363-378, 1991.
- [36] V. R. Fuchs and I. Leveson. Motor accident mortality and compulsory inpection of vehicles. Journal of the American Medical Association, 201:87-91, 1967.
- [37] C. Garbacz. A note on pelzmann's offsetting consumer behaviour. Economic Letters, 19:183-7, 1985.
- [38] Christopher Garbacz and J. Gregory Kelly. Automobile safety inpection: new econometric and benefit/cost estimates. Applied Economics, 19:763-771, 1987.
- [39] Stephen Godwin. International experience with speed limits during and prior to the energy crisis of 1973-4. Transport Planning and Technology, 9(1):25-36, 1984.
- [40] C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: application to poisson models. *Econometrica*, 52:701-720, 1984.
- [41] J. D. Graham and S. Garber. Evaluating the effects of automobile safety regulations. Journal of Policy Analysis and Management, 3:206-224, 1984.
- [42] Expert group on road safety. Evaluation of countermeasures. In Road accident information seminar (Canberra 26-28 March 1974 — Proceedings), AGPS, 1978.
- [43] F.A. Haight. Traffic safety in developing countries. Journal of Safety Research, 12(2):50-58, 1980.

- [44] Simon Hakim, Daniel Shefer, A. S. Hakkert, and Irit Hocherman. A critical review of macro models for road accidents. Accident Analysis and Prevention, 23(5):379-400, 1991.
- [45] J. Hall and L. Dickinson. Truck speeds and accidents on interstate highways. Transportation Research Record, 486:19-32, 1974.
- [46] G. Hampson. The theory of accident compensation and the introduction of compulsory seatbelt legislation in NSW. In Proceedings ARRB Conference, pages 135-140, 1982.
- [47] A. C. Harvey and J. Durbin. The effects of seat belt legislation on british road casualties: a case study in structural time series modelling (with discussion.). Journal of the Royal Statistical Society, A: 149:187-227, 1986.
- [48] E. Hauer. Accidents, overtaking and speed control. Accident Analysis and Prevention, 3, 1971.
- [49] E. Hauer. Bias-by-selection:overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. Accident Analysis and Prevention, 12:113-117, 1980a.
- [50] E. Hauer. On the estimation of the expected number of accidents. Accident Analysis and Prevention, 18(1):1-12, February 1986.
- [51] Heinz Hautzinger. Regression analysis of aggregate accident data: some methodological considerations and practical experiences. Accident Analysis and Prevention, 18(2):95-102, 1986.
- [52] J. D. Havard. Mortality from motor vehicle accidents in the 15-24 year age group. World Health Statistics Quarterly, 32:225-241, 1979.
- [53] James Hedlund. Comments on Hauer's approach to statistical inference. Accident Analysis and Prevention, 16:163-165, 1984.
- [54] Ulrich Helfenstein. When did a reduced speed limit show an effect? Exploratory identification of an intervention time. Accident Analysis and Prevention, 22(1):79-87, 1990.
- [55] Ulrike Henning-Hager. Urban development and road safety. Accident Analysis and Prevention, 18(2):135-145, 1986.
- [56] T. Hutchinson. Road accident statistics. SA Rumsby Scientific Publ. Co., Adelaide, 1987.
- [57] D. Ivey, L. I. Griffin, T. M. Newton, R. L. Lytton, and K. C. Hankins. Predicting wet weather accidents. Accident Analysis and Prevention, 13:83-99, 1981.

- [58] G. D. Jacobs and C. A. Cutting. Further research on accident rates in developing countries. Accident Analysis and Prevention, 18(2):119-127, 1986.
- [59] G. D. Jacobs and W. Hards. Further research on road accidents in developing countries. TRRL Report SR 434, Transport and Road Research Laboratory, 1977.
- [60] G. D. Jacobs and I. Sayer. Road accidents in developing countries. 1983.
- [61] K. S. Jadaan and R. J. Salter. Traffic accidents in Kuwait. Traffic Engineering and Control, 23(4):221-223, April 1982.
- [62] Mary K. Janke. Accidents, mileage and the exaggeration of risk. Accident Analysis and Prevention, 23:183-188, 1991.
- [63] H.D. Johnson and F. Garwood. Notes on road accident statistics. RRL Report, Road Research Laboratory, 1971.
- [64] P. Johnson, T. M. Klein, P. Levy, and D. Maxwell. The effectiveness of the 55 mph national maximum speed limit as a life saving benefit. Tech report HS-805 694, National Highway Traffic Safety Administration, Oct 1980.
- [65] S. R. Johnson and J. D. McQuigg. Some useful applications to the measurement of economic relationships which include climatic variables, chapter 12. Newton Abbott, 1974.
- [66] H. C. Joksch. Are traffic deaths in Ireland related to alcohol consumption? Restudying a time series model. Accident Analysis and Prevention, 23(5):437-442, 1991.
- [67] H. C. Joksch. Critique of Sam Peltzmann's study: the effects of automobile regulation. Accident Analysis and Prevention, 8(2):129-137, 1976.
- [68] H. C. Joksch. The relation between motor vehicle accident deaths and economic activity. Accident Analysis and Prevention, 16:207-210, 1984.
- [69] H. C. Joksch and H. Wuerdermann. Estimating the effects of crash phase injury countermeasures II. the fatality trend and its modification by countermeasures. Accident Analysis and Prevention, 5:1-26, 1973.
- [70] P. Jovanis and J. Delleur. Exposure-based analysis of motor vehicle accident. Transportation Research Record, 910:1-7, 1983.
- [71] Paul P. Jovanis and Hsin-Li Chang. Disaggregate model of highway accident occurrence using survival theory. Accident Analysis and Prevention, 21(5):445-458, 1989.

- [72] P.P. Jovanis and H.-L. Chang. Modeling the relationship of accidents to miles travelled. Transportation Research Record, 1068:42-51, 1986.
- [73] Rajindar K. Koshal. Deaths from road accidents in the United States: an econometric analysis. Journal of Transport Economics and Policy, September:219-226, 1976.
- [74] S. Lassarre. The introduction of the variables traffic volume, speed and belt-wearing into a predictive model of the severity of accidents. Accident Analysis & Prevention, 18(2):129-34, April 1986.
- [75] S. Lassarre and S. H. Tan. Evaluation of safety measures on the frequency and gravity of traffic accidents in France by means of intervention analysis., chapter 1, pages 10-28. North Holland, Amsterdam, 1982.
- [76] Charles A. Lave. Speeding, coordination, and the 55 mph limit. The American Economic Review, 1159-1164, December 1985.
- [77] R. Lavette. Development and application of railroad-highway accident prediction equation. Transportation Research Record, 628:12-19, 1977.
- [78] K. S. Lindqvist. Epidemiology of accidents in a Swedish municipality. Accident Analysis and Prevention, 21(1):33-43, February 1989.
- [79] P. D. Loeb and B. Gilad. The efficacy and cost-effectiveness of vehicle inspection: a state specific analysis using time series data. Journal of Transport Economics and Policy, 18(2):145-64, May 1984.
- [80] Peter D. Loeb. The determinants of automobile fatalities. Journal of Transport Economics and Policy, 279-287, September 1987.
- [81] Angus MacKenzie. Road safety: licence to kill. The Bulletin, November 1990.
- [82] A. S. MacLean and G. Teale. Probability distributions of traffic accidents, injuries and deaths. ARRB Journal, 22(1):42-44, March 1982.
- [83] Scott MacLean. Relationships between variables. In Esso-Monash Workshop on Traffic Accident Evaluation: Feb 15-17, 1983.
- [84] Mahalel. A note on accident risk. Transportation Research Record, 1068:85-89, 1986.
- [85] Robert E. Mann and Lise Anglin. The relationship between alcoholrelated traffic fatalities and per capita consumption of alcohol, Ontario, 1957-1983. Accident Analysis and Prevention, 20(6):441-446, 1988.
- [86] Beverly Martinez-Schnell and Akbar Zaidi. Time-series analysis of injuries. Statistics in Medicine, 8:1497-1508, 1989.

a
- [87] W. J. Maunder. National econoclimatic models, chapter 13. Newton Abbott, 1974.
- [88] A. J. Mayer and T. F. Hoult. Motor vehicle inspection: a report on current information, measurement and research. Technical Report, Institute for Regional and Urban studies, Wayne State University, 1963.
- [89] P McCullagh and J. A Nelder. Generalised Linear Models. London: Chapman and Hall, 1983
- [90] A. Mekky. Effects of rapid increase in motorization levels on road fatality rates in some rich developing countries. Accident Analysis and Prevention, 17(2):101-9, April 1985.
- [91] G. William Mercer. The relationships among driving while impaired charges, police dring-driving roadcheck activity, media coverage and alcohol-related casualty traffic accidents. Accident Analysis and Prevention, 17(2):467-474, 1985.
- [92] A. L. Minter. Road casualties—improvement by learning processes. Traffic Engineering and Control, 28(2):74-79, February 1987.
- [93] H. Moskowitz, G. Yourgrau, and B. Wurf. Alcohol, drugs & driving, abstracts & reviews. Alcohol, Drugs & Driving, 2(1), January 1986.
- [94] Takahiro Murata. Japan's struggle for traffic accident reduction. Transportation Research 4, 23(1):83-90, 1989.
- [95] T. R. Neuman. Discriminant analysis as a tool for investigating highway safety relationships. *ITE Journal*, 54:16-22, October 1984.
- [96] R. H. Oldfield. A theoretical model for estimating the effects of fares, traffic restraint and bus priority in central london. Transport and Road Research Laboratory Lub Report, 749:0305-1293, 1977.
- [97] S. Oppe. The development of traffic and traffic safety in six developed countries. Accident Analysis and Prevention, 23(5):401-412, 1991.
- [98] S. Oppe. Macroscopic models for traffic and traffic safety. Accident Analysis and Prevention, 21(3) 225-32, June 1989.
- [99] S. Oppe. The use of multiplicative models for analysis of road safety data. Accident Analysis and Prevention, 11:101, 1979.
- [100] Susan C. Partyka. Simple models of fatality trends revisited seven years later. Accident Analysis and Prevention, 23(5):423-430, 1991.
- [101] Susan C. Partyka. Simple models of fatality trends using employment and population data. Accident Analysis and Prevention, 16:211-222, 1984.

- [102] Sam Peltzmann. The effects of automobile safety regulation. Journal of Political Economy, 83(4):677-725, 1975.
- [103] Sam Peltzmann. The effects of automobile safety regulations: reply. Accident Analysis and Prevention, 8:139-142, 1976.
- [104] Bhagwant Persaud. Relating the effect of safety measures to expected number of accidents. Accident Analysis and Prevention, 18(1):63-70, 1986.
- [105] B. Preston. Road safety: international comparisons. Transportation Reviews, 1(1):75-98, 1982.
- [106] Donald W. Reinfurt, J. Richard Stewart, and Nancy L. Weaver. The economy as a factor in motor vehicle fatalities, suicides and homocides. Accident Analysis and Prevention, 23(5):453-462, 1991.
- [107] Leon S. Robertson. A critical analysis of Peltzmann's 'The effects of automobile safety regulation'. Journal of Economic Issues, 11(3), September 1977.
- [108] P. P. Scott. Modelling time-series of British road accident data. Accident Analysis and Prevention, 18(2), April 1986.
- [109] P. P. Scott. Variations in two-vehicle accident frequencies, 1970-78. Report LR1086, Transport and Road Research Laboratory, 1983.
- [110] M. Sivak. Society's aggression level as a predictor of the traffic fatality rate. Journal of Safety Research, 14:93-99, 1983.
- [111] R. J. Smeed. Some statistical aspects of road safety. Journal Royal Statistical Society A, 102(4):1-24, 1949.
- [112] R. J. Smeed and G. O. Jeffcoate. Effect of changes in motorisation in various countries on the number of road fatalities. Traffic Engineering and Control, 12:150-151, 1970.
- [113] Geraldine Spencer. Australian mortality: a study of causes of death. 1980.
- [114] J. C. Tanner. Casualty rates at Easter, Whitsunday, and August Public Holidays. Laboratory report LR 74, Ministry of Transport, Road Research Laboratory, 1967.
- [115] J. C. Tanner. Effect of weather on traffic flow. Nature, 169(4290):107, 1952.
- [116] J. C. Tanner. Weather and road traffic flow. Weather, 7(9):270-5, 1952.

- [117] J. Thomson. Road loss research in Australia: the potential contribution of econometrics. In Proceedings of 11th Australian Road Research Board Conference, University Melbourne, pages 221-232, Australian Road Research Board, August 1982.
- [118] J. et al Treat. Tri-level study of the causes of traffic accidents. Technical Report, Springfield VA: National Technical Information Service, 1977.
- [119] H. L. Votey. Taking account of system interactions in modelling road accidents. Accident Analysis and Prevention, 18(2):85-94, April 1986.
- [120] Alexander C. Wagenaar. Effects of macroeconomic conditions on the incidence of motor vehicle accidents. Accident Analysis & Prevention, 16:191-205, 1984.
- [121] Alexander C. Wagenaar, Federic M. Streff, and Robert H. Schultz. Effects of the 65 mph speed limit on injury morbidity and mortality. Accident Analysis & Prevention, 22(6):571-585, 1990.
- [122] Brendan M. Walsh. Do excise taxes save lives? The Irish experience with alcohol taxation. Accident Analysis and Prevention, 19:433-448, 1987.
- [123] P. Wasielewski. Speed as a measure of driver risk: observed speed versus driver and vehicle characteristics. Accident Analysis and Prevention, 16:89-103, 1984.
- [124] P. Wasielewski and L. Evans. A statistical approach to esimating driver responsibility in two-car crashes. Journal of Safety Research, 16:37-48, 1985.
- [125] Richard M. Weed. Revised decision criteria for before-and-after analyses. Transportation Research Record, 1068:8-17, 1986.
- [126] George H. Weiss. An appreciation of the work of Reuben J. Smeed. Transportation Research A, 19(2):85-88, 1985.
- [127] W. C. Wheaton. The long run structure of transportation and gasoline demand. Bell Journal of Economy, 13:439-454, 1982.
- [128] W. T. White. Does periodic vehicle inspection prevent accidents. Accident Analysis and Prevention, 18(1):51-62, 1986.
- [129] W.H.O. World statistics. 1982.
- [130] G. J. Wintemute. Is motor vehicle-related mortality a disease of development. Accident Analysis and Prevention, 17(3):223-7, June 1985.
- [131] G. J. Wintemute. Motor-vehicle related mortality and economic development. American Association Automotive Medical Journal, 6, January 1984.

- [132] J. Zlatoper, Thomas. Regression analysis of time series data on motor vehicle deaths in the United States. Journal of Transport and Economic Policy, 18:263-274, September 1984.
- [133] T. J. Zlatoper. Models explaining motor vehicle death rates in the US. Accident Analysis and Prevention, 21(2):125-54, April 1989.
- [134] Thomas J. Zlatoper. Factors affecting motor vehicle deaths in the USA: some cross-sectional evidence. Applied Economics, 19:753-761, 1987.

## Part 2

.

# Explanatory and Predictive Models for the Number of Fatal Road Crashes

#### © Statistical Consulting Unit, QUT, 1992

This report has been prepared specifically for the Department of Transport & Communication as the client. Neither the report nor its contents may be referred to or quoted in any statement, study, report, application, prospectus, loan, other agreement or document, without the express approval of the Statistical Consulting Unit, QUT.

The information contained in this report is based on sources believed to be reliable. However, as no independent verification is possible, the QUT together with its members and employees gives no warranty that the said base sources are correct, and accepts no responsibility for any resultant errors contained herein and any damage or loss, howsoever caused, suffered by any individual or corporation.

## Glossary

PART 1	TIME SERIES MODELS
	§1 Introduction
	§2 Results for Australian Monthly Series
	§3 Results for State Monthly Series
	§4 Comparison of States and Australia
PART 2	MODELS INVOLVING EXPLANATORY
	VARIABLES
	§5 General Methodology
	§6 Results for Monthly Series
	§7 Results for Quarterly Series
	§8 Explanatory & Predictive Models for Routine Use
PART 3	FATALATIES PER CRASH
	§9 Modelling of the Number of Fatalities per Crash
REFERENCES	
APPENDICES	Appendix A
	Tables and Figures for Monthly and Quarterly series of number
	of fatal road crashes and explanatory variables
	Appendix B
	Sources of explanatory variables
	Appendix C
	Computer code for predictive models
	Appendix D
	Figures for fatalaties per crash analyses

## Glossary

ī.

NCF	Number of crash fatalities.
NFRC	Number of fatal road crashes.
FC	NCF divided by NFRC.
AUSACCM	NFRC monthly series for Australia standardised by days in the month, i.e. average daily NFRC for a given month for Australia.
AUSACCQ	NFRC quarterly series for Australia standardised by estimated population of Australia.
NSWACCM	NFRC monthly series for New South Wales standardised by days in the month, i.e. average daily NFRC for a given month for New South Wales.
NSWACCQ	NFRC quarterly series for New South Wales standardised by estimated population of New South Wales.
QLDACCM	NFRC monthly series for Queensland standardised by days in the month, i.e. average daily NFRC for a given month for Queensland.
QLDACCQ	NFRC quarterly series for Queensland standardised by estimated population of Queensland.
SAACCM	NFRC monthly series for South Australia standardised by days in the month, i.e. average daily NFRC for a given month for South Australia.
TASACCM	NFRC monthly series for Tasmania standardised by days in the month, i.e. average daily NFRC for a given month for Tasmania.
VICACCM	NFRC monthly series for Victoria standardised by days in the month, i.e. average daily NFRC for a given month for Victoria.
VICACCQ	NFRC quarterly series for Victoria standardised by estimated population of Victoria.
WAACCM	NFRC monthly series for Western Australia standardised by days in the month, i.e. average daily NFRC for a given month for Western Australia.

,

## **Glossary** of Symbols

В	backshift operator, $By_t = y_{t-1}$
B <sup>4</sup>	annual backshift operator for quarterly data, $B^4y_t = y_{t-4}$
B <sup>12</sup>	annual backshift operator for monthly data, $B^{12}y_t = y_{t-4}$
$\nabla$	first difference operator, $\nabla y_t = y_t - y_{t-1} \nabla = 1 - B$
$\nabla_4$	annual difference operator for annual data, $\nabla_4 y_t = y_t - y_{t-4} \nabla_4 = 1 - B^4$
$\nabla_{12}$	annual difference opeator, $\nabla_{12}y_t = y_t - y_{t-12}$
α <sub>t</sub>	level parameter for trend
β <sub>t</sub>	regression parameter for independent variable
$\delta_t$	slope parameter for trend
μ <sub>t</sub>	trend parameter
γt	seasonal parameter
e <sub>t</sub>	noise parameter
x <sub>t</sub>	independent explanatory variable

yt dependent response, time series data

## PART 1

## **Time Series Analysis**

#### Overview

ARIMA time series models are fitted to the monthly number of fatal road crashes series for Australia and the states. For all series, it is found that the so-called 'airline' model provides a good fit to the data. For Australia and NSW, the data for the period March 1983 to December 1990 is used, whereas for other states, the period January 1976 until December 1990 is used. For the Australia and NSW series there is an apparent discontinuity in average values around February 1983. There is a large amount of similarity between estimated parameter values for the 'airline' models fitted to the Australia and states series. The models are used to predict values for January 1991 to June 1991 using data up to December 1990 as a base. Predictions are compared with actual values. Prediction errors are consistent with the inherent variability which is to be expected to be found in monthly counts such as the number of fatal accidents. That is, the prediction model is performing as well as any prediction model could perform. However, there is a consistent overestimate of fatal road crashes for June 1991 across all states except Tasmania.

#### 1. INTRODUCTION

#### 1.1 Background

In developing a time series model for NFRC series, there are two basic approaches:

- . a model with explicit trend and seasonal components;
- a model with implicit trend and seasonal components defined by autoregressive linear models and correlated error structures, an example of which is the Box-Jenkins ARIMA approach (Box and Jenkins, 1970).

The Box-Jenkins approach is adopted in this part of the study while the explicit trend and seasonal components approach is applied in the section on explanatory variable modelling. The Box-Jenkins approach has been used extensively in previous studies on accident data. Section 7 of volume 1 of this report mentions the Australian study by Bhattacharrya *et al* (1979), an early study using the Box-Jenkins approach. More recently Harvey and Durbin (1986) use structural modes for UK data but for prediction purposes use a Box-Jenkins type model.

#### 1.2 Standardisation

A characteristic of ARIMA modelling is that after applying varying amounts of differencing the resultant series is assumed to have a constant mean. For monthly NFRC series, we can assume that

mean = number days in month x daily rate

$$E(y_t) = n_t \lambda_t$$

Suggesting that monthly count data should be standardised by the number of days in the month, giving

$$E\left(\frac{y_t}{n_t}\right) = \lambda_{t'}$$

To ensure that known monthly effects are removed from the series, we have conducted the time series analysis of the NFRC series using the standardised (by days in the month) series. For both additive and multiplicative models, the effect of the varying number of days in the month would be accommodated by seasonal effects but we believe it is wiser to remove this calendar effect first.

If the daily number of accidents is assumed to follow a Poisson distribution, then the variance of the daily rate for a month is given by

$$\operatorname{var}\left(\frac{\operatorname{y}_{\operatorname{t}}}{\operatorname{n}_{\operatorname{t}}}\right) = \frac{\lambda_{\operatorname{t}}}{\operatorname{n}_{\operatorname{t}}}$$

 $var(y_t) = n_t \lambda_t$ 

whereas

For the models fitted here we assume that  $var(y_t/n_t)$  is well approximated by a constant, that is fitting procedures assume constant variance.

Standardisation by a variable dependent on the calendar is standard practice in economics for example, where trading days per calendar month would be an appropriate divisor. Dividing by a variable such as days in the month, which are known precisely, is not problematic; dividing by a variable which is measured with error is problematic.

#### **1.3 Additive or Multiplicative Models**

With the NFRC series there is a choice between modelling the series on the original scale or on the log scale. The former leads to an additive model, whilst the latter leads to a multiplicative model. If the signal to noise ratio is small, then, typically, there will be little difference between the two approaches. Additionally, if the range of the values of the series is small, say max/min < 2, then there will be little difference, generally, between fitted and predicted values for additive and multiplicative models.

Generally, studies undertaken and reported in Volume 1 tend to model the variation although on the log scale. Certainly when converting predictions on the log scale to predictions on the original scale there are problems of unbiasedness, etc. We prefer to model on the original scale and avoid these problems but alternative analyses on the log scale have been undertaken for comparative purposes. There is little difference as suggested above.

#### **1.4 ARIMA Models with Accident Data**

Typically for ARIMA modelling of series showing trend and seasonal patterns the socalled 'airline' model (for monthly data)

$$(1 - B)(1 - B^{12}) y_t = (1 - \theta B)(1 - \Theta B^{12}) e_t$$

provides the best model amongst the class of ARIMA models; see Box and Jenkins (1976, § 9). Here  $y_t$  is the (standardised - that is the average number of crashes per day in a given month) monthly series, B is the backshift operator. By<sub>t</sub> =  $y_{t-1}$ ,  $B^{12}y_t$  =

 $y_{t-12}$ ,  $\theta$ ,  $\Theta$  are parameters to be estimated, and  $e_t$  is white noise. Harvey and Durbin (1986) fitted the 'airline' model to accident data (car drivers killed and seriously injured) from the U.K. for the period 1969 - 1985 and estimated

$$\hat{\theta} = +0.684, \ \hat{\Theta} = +0.995$$

using the subseries 1969 - 1982 (note change of sign from their non-standard formulation to ours). The operator '1 - B' removes trend from the series and the operator '1 -  $B^{12}$ ' removes the seasonal component.

With  $\Theta = 1$ , the operator  $(1 - B^{12})$  cancels on both sides leaving a simpler model

$$(1 - B) y_t = \gamma_t + (1 - \theta B) e_t$$

involving only first differences, an autocorrelated error structure and fixed monthly effects,  $\gamma_t = \gamma_{t-12}$ .

For quarterly data the operator  $B^{12}$  is replaced by  $B^4$  with  $B^4y_t = y_{t-4}$ .

The statistical software package STATGRAPHICS (Statistical Graphics Corporation, 1989) was used for fitting ARIMA models to the data. The package provides estimated

values of the parameters  $\theta$  and  $\Theta$ , denoted by MA (1) and SMA (12), respectively, in the package and in what follows. Diagnostic statistics for the adequacy of the fitted model include the following

- a portmanteau goodness-of-fit statistic being the weighted sum of squares of the terms of the residual (one-step ahead prediction error) autocorrelation function (acf) and having an asymptotic chi-squared distribution given white noise for the residual error series (see Box and Pierce, 1970). A modified Box-Pierce statistic (Ljung and Box, 1978) has its distribution approximated better by the chi-squared distribution for small samples. For the sample size n = 180 used mainly in these studies, the correction suggested by Ljung and Box has only a small effect but for sample sizes less than 100 (also used in the study) the values of the chi-squared statistic printed by STATGRAPHICS should be inflated by 10 percent for a reasonable correction. The statistic given in the Tables is the Box-Pierce statistic for n = 180 and the Box Pierce statistic inflated by 10 percent for n < 100. The statistic is based on the first 20 autocorrelations.
- measures of skewness and kurtosis for the residuals.

These are supplemented by plots of the acf and partial acf against lag, and plots of ordered residuals against expected normal percentiles (Q - Q plots). These plots are not given in this report.

#### 2. RESULTS FOR AUSTRALIAN MONTHLY SERIES

#### 2.1 Estimated Model

A plot of the daily average NFRC for the monthly series for Australia (AUSACCM) for 1976 until 1990 is given in Appendix A, Figure A.1 and shows a steady decline overall with a relatively sharp decline in early 1983.

An ARIMA model was fitted to three series of the AUSACCM data to investigate the stability of parameters over time:

Jan	$1976 - \text{Dec } 1990 \ (n = 180)$
Jan	1976 - Feb  1983 (n = 86)
March	1983 - Dec  1990 (n = 94)

Summaries of estimates and diagnostic statistics are given in Tables 2.1a, 2.1b and 2.1c.

In each case the 'airline' model gave a good fit to the data with estimates in the range 0.74 to 0.84 for MA(1) and 0.66 to 0.73 for SMA(12) for the three series. Residual analyses showed no unusual features. The noise standard deviation was smaller for the later series, March 1983 - December 1990 than the earlier series, by about 20 percent. Typically, a value of AUSACCM in 1990 is about 5.7 fatal accidents per day, so that a noise standard deviation of about 0.7 gives a relative error of about 0.7/5.7 or 12 percent. If the parameters were known precisely this would be the one step ahead relative error of prediction for the fitted model.

#### 2.2 Predictions for 1991

In Table 2.1d we give predictions and confidence intervals for values of the series for 1991 based on the model fitted to the March 1983 - December 1990 series.

Predictions for the first months of 1991 from data up to December 1990 are reasonable except that the June value is overestimated by 0.98 accidents per day. All actual values are within the 95% confidence limits. The 95% confidence intervals are wide; typically the upper value is almost twice the lower value. The six errors of prediction, actual value minus predicted value, are both negative and positive with an average value of - 0.04 accidents per day, indicating little bias of prediction.

Multiplicative models were fitted to the data by taking logarithms of the standardised series. Little difference was found between fitted values and predictions for an additive model and a multiplicative model. On the grounds of simplicity, an additive model was used.

In Table 2.1e we give prediction errors for data obtained in January 1992 from FORS. For the year 1991 there are six positive and six negative errors and the average absolute error is 0.39.

#### Table 2.1a

#### AUSACCM : Estimates for 'airline' ARIMA model

	Jan estimate	1976 - Dec std error	1990 t value
MA (1)	0.804	0.046	17.4
SMA (12)	0.734	0.057	12.8
Noise std dev	0.72		
Residuals			
skewness	-0.079		-0.42
kurtosis	-0.63		-1.65
chi squared statistic = $18.12 (p = 0.45)$			

#### Table 2.1b

#### AUSACCM : Estimates for 'airline' ARIMA model

	Jan 1976 - Feb 1983		
	estimate	std error	t value
MA (1)	0.84	0.066	12.8
SMA (12)	0.66	0.099	6.66
Noise std devn	0.83		
Residuals			
skewness	-0.093		-0.33
kurtosis	-0.89		-1.55
chi squared statistic = $13.00 (p = 0.79)$			

#### Table 2.1c

	Ma	ec 1990	
	estimate	std error	t value
MA (1)	0.74	0.075	9.95
SMA (12)	0.68	0.098	6.89
Noise std devn	0.67		
Residuals			
skewness	0.19		0.70
kurtosis	-0.59		-1.09
chi squared statistics = 20.6	s(p = 0.30)		

#### AUSACCM : Estimates for 'airline' ARIMA model

#### Table 2.1d

#### AUSACCM : Predicted values & 95% confidence limits for 1991 using model in Table 2.1c

month	predicted	confidence	limits	actual	error
January	4.63	3.29	5.97	158/31 = 5.10	0.47
February	4.87	3.50	6.27	139/28 = 4.96	0.09
March	5.70	4.27	7.11	171/31 = 5.52	- 0.18
April	4.63	3.16	6.10	152/30 = 5.07	0.44
May	4,91	3.40	6.41	154/31 = 4.97	0.06
June	5.31	3.77	6.85	130/30 = 4.33	- 0.98
July	5.02	3.44	6.60	····	
August	4.77	3.15	6.38		
September	5.92	4.27	7.57		
October	4.91	3.23	6.60		
November	5.03	3.31	6.75		
December	5.76	4.01	7.52		

#### Table 2.1e

#### AUSACCM : Predicted values & 95% confidence limits for 1991 using model in Table 2.1c

month	predicted	confidence	limits	actual*	error
January	4.63	3.29	5.97	157/31 = 5.06	0.43
February	4.87	3.50	6.27	137/28 = 4.89	0.02
March	5.70	4.27	7.11	170/31 = 5.48	- 0.22
April	4.63	3.16	6.10	156/30 = 5.20	0.53
May	4.91	3.40	6.41	150/31 = 4.84	0.07
June	5.31	3.77	6.85	138/30 = 4.60	- 0.71
July	5.02	3.44	6.60	163/31 = 5.26	0.24
August	4.77	3.15	6.38	171.31 = 5.52	0.75
September	5.92	4.27	7.57	161/30 = 5.37	- 0.55
October	4.91	3.23	6.60	166/31 = 5.35	0.44
November	5.03	3.31	6.75	148/30 = 4.93	- 0.10
December	5.76	4.01	7.52	159/31 = 5.13	- 0.67

average lerror = 0.34

\* obtained January 1992 from FORS

#### 3. RESULTS FOR STATE MONTHLY SERIES

#### 3.1 Introduction

The ARIMA modelling approach was used to investigate the NFRC monthly series for each state. The territories were omitted because monthly rates are too small for analysis by the ARIMA models used here; similar comments hold for Tasmania. In each case the 'airline' model was found to be a good model with parameter estimates similar to those for the Australian series. Again, the period 1976 - 1990 was considered for splitting into two periods. The disaggregated state series have distinctive differences but are all well modelled by very similar ARIMA 'airline' models.

#### 3.2 NSW Monthly Data

Appendix A Figure A.3 gives the NSWACCM data plotted, and the plot shows a distinct change in level near February 1983. Results for the three series, the full period 1976 - 1990, before and after 1983 are given in Tables 3.1a, 3.1b, 3.1c. Estimates of MA (1) and SMA (12) are similar, close to 0.8 and 0.7 respectively, to those for the AUSACCM series. For the period 1976 - 1990, the noise standard deviation is 0.43, whereas for March 1983 - December 1990, the noise standard deviation is 0.38. This indicates better predictions for the model fitted to 1983 - 1990 data. The average for the values in 1990 is about 2.0 fatal crashes per day, giving a relative error of about 20 percent.

Predictions are given in Table 3.1d based on the model fitted to the March 1983 - December 1990 data.

The predictions for the first months of 1991 are reasonable except for the over estimate for March. The errors of prediction have an average value of -0.10 accidents per day. All actual values are within the 95% confidence intervals.

#### 3.3 Queensland Monthly Data

A plot of the Queensland series, QLDACCM, in Appendix A Figure A.8, shows no abrupt drop in values which the series for NSW shows in 1983. Consequently, only the complete series, 1976 - 1990, was analysed. Parameter estimates were not significantly different from 0.8, for MA (1) and not significantly different from 0.7 for SMA (12); Table 3.2a. The noise standard deviation is 0.25 whilst the 1990 value of the series is typically close to 1.0 giving a 25 percent relative error.

Predictions for 1991 based on the model fitted to the January 1976 to December 1990 values of the series are given in Table 3.2b. Predictions are reasonable with June 1991 being under estimated by an amount 0.38 accidents per day. The average error of prediction is 0.08 accidents per day while 3 of the errors of prediction are negative and 3 positive. All actual values are within the 95% confidence intervals, which are wide.

#### 3.4 South Australia Monthly Data

Like the QLD series there was no abrupt change in values about 1983 for the SA series and the level is relatively constant over the period of the series (Appendix A, Figure A.10). From Table 3.3a, the parameter estimate, 0.67, for MA (1), is significantly different (t value = 2.2) from the value 0.8 obtained for the Australian series but only just; the estimate for SMA (12) is not significantly different from the value 0.7 obtained for the Australian series. The noise standard deviation is 0.21 compared with a typical value of about 0.4 for value in 1990, giving a 50 percent relative (prediction) error. Predictions for 1991 based on the model fitted to the January 1976 to December 1990 values of the series are given in Table 3.3b. There appears to be some under estimate of values with the average error of prediction being 0.06 accidents per day. Both February and March 1991 are under estimated so that predictions are about 60% of actual values. The lower values of the confidence intervals for predictions are in each case negative and this has been replaced by zero. The original confidence intervals are symmetric about the predicted value and do not take into account the necessary non-negativeness of the predictions. For SA, the predictions are not particularly good but we comment on this later.

#### 3.5 Tasmania Monthly Data

The TASACCM series shows no abrupt change in values in 1983 and only one series (Jan 1976 - Dec 1990) was analysed (Appendix A, Figure A.10). From Table 3.4a the MA (1) parameter estimate of 0.92 is significantly different from the value 0.8 (t value = 3.75); the SMA (12) estimate is not significantly different from the value 0.7. The noise standard deviation is 0.10 and the average value of the series for 1990 is about 0.17, giving a relative error of almost 60 percent.

The kurtosis of the residuals is a little large and positive but a Q - Q plot of residuals is close to being linear required for normally distributed errors.

Predictions for 1991 based on the model fitted to the January 1976 to December 1990 data are given in Table 3.4b. Conficence intervals are wide, the upper limit being about twice the predicted value. Values for Tasmania are generally between 5 and 10 accidents per month and so the appropriateness of the ARIMA approach can be questioned. Nevertheless, the 'airline' model gives predictions and these tend to be too small, giving underestimates. Results are included for completeness.

#### 3.6 Victoria Monthly Data

The Victorian series, VICACCM (Appendix A, Figure A.5, Figure A.10), shows a substantial change in level at the end of 1979 and beginning of 1980. The full series Jan 1976 - Dec 1990 was analysed and subseries were considered with little difference between results, so that the full series was used to estimate parameters. From Table 3.5a the values of the parameters are very close to the values 0.8 and 0.7 and the 'airline' model appears to fit well. The noise standard deviation is 0.29 compared with an average 1990 value close to 1.4 giving a relative error of about 20 percent.

Predictions for 1991, based on the model fitted to the January 1976 to December 1990 data, are given in Table 3.5b. Predictions, actual values and errors of prediction give an interesting pattern. The values for the first three months, January, February and March are underestimated by a small amount, about 0.10 accidents per day, whereas predictions for the next three months, April, May and June are overestimates by about 0.43 accidents per day or, in relative terms, by about 50 percent. This suggests that there has been a dramatic change in the underlying process from April 1991 onwards, and that the database up to December 1990 is not adequate to predict the changes seen from April 1991 onwards.

#### 3.7 Western Australia Monthly Data

The series for WA appears to be somewhat different from that for the eastern states (Appendix A, Figure A.10). There is no reduction in values towards the end of 1979 and beginning of 1980 as experienced in Victoria; in fact the general level appears to increase. There is no abrupt reduction in level in early 1983 as seen in the NSW series;

in fact again the level appears to increase. However, from Table 3.6a, the MA (1) and SMA (12) estimates are not significantly different from the values 0.8 and 0.7. The noise standard deviation is 0.17 while the average value of the series for 1990 is about 0.5 giving a relative error of 34 percent. The chi-squared statistic, computed from the first terms of the residual autocorrelation function (acf), has a p-value of 0.06. However, on inspection of the residual acf, there is no single value significantly different from 0 at the 5 percent level. Hence, no alternative model is suggested by the residual acf.

Predictions for 1991, based on the model fitted to the January 1976 to December 1990 data, are given in Table 3.6b. It should be noted first that predictions for the early months of 1991 are very similar at 0.50 accidents per day. There is a small deviation from this value for the predictions for May and June 1991. A large underestimate, equal to 0.25 accidents per day, or, in relative terms, about 30 percent, occurs for January 1991. For the remaining months there tends to be an overestimate. The average error of prediction for the six months is - 0.03 accidents per day, that is an overestimate on average. All actual values lie in the 95 percent confidence intervals which tend to be wide.

#### Table 3.1a

#### NSWACCM : Estimates for 'airline' model

	January 1976 - Dec 1990			
	estimate	std error	t value	
MA (1)	0.84	0.043	19.4	
SMA (12)	0.74	0.056	13.3	
Noise std dev	0.43			
Residuals				
skewness	-0.094		-0.49	
kurtosis	-0.49		-1.29	
chi squared statistic = $24.7 (p = 0.13)$				

#### Table 3.1b

#### NSWACCM : Estimates for 'airline' model

	January 1976 -			
	estimate	std error	t value	
MA (1)	0.78	0.089	8.8	
SMA (12)	0.70	0.097	7.2	
Noise std dev	0.51			
Residuals				
skewness	-0.12		-0.42	
kurtosis chi squared statistic = $17.4 (p = 0.50)$	-0.89		-1.56	

#### Table 3.1c

	Ma	1990	
	estimate	std error	t value
MA (1)	0.84	0.06	13.5
SMA (12)	0.67	0.10	6.5
Noise std devn	0.38		
Residuals			
skewness	-0.34		-1.25
kurtosis	-0.56		-1.02
chi squared statistic = $23.3 (p = 0.18)$			

#### NSWACCM : Estimates for 'airline' model

#### Table 3.1d

#### NSWACCM : Predicted values & 95% confidence intervals for 1991 using Model of Table 3.1c

month	predicted	confidence	limits	actual	error
January	1.46	0.69	2.21	58/31 = 1.87	0.41
February	1.76	0.99	2.53	44/28 = 1.57	- 0.19
March	2.06	1.28	2.84	42/31 = 1.35	- 0.71
April	1.49	0.70	2.29	50/30 = 1.67	0.18
May	1.66	0.86	2.46	53/31 = 1.71	0.05
June	1.83	1.02	2.64	45/30 = 1.50	- 0.33
July	1.87	1.05	2.69		
August	1.55	0.72	2.38		
September	2.11	1.27	2.94		
October	1.89	1.04	2.73		
November	1.68	0.83	2.53		
December	1.98	1.12	2.84		

#### Table 3.2a

#### QLDACCM : Estimates for 'airline' ARIMA model

	Jan 1976 - Dec 1990		
	estimate	std error	t value
MA (1)	0.72	0.052	13.9
SMA (12)	0.73	0.056	13.1
Noise std devn	0.25		
Residuals			
skewness	-0.021		-0.11
kurtosis chi squared statistic = $16.07 (p = 0.59)$	-0.26		-0.70

.

#### Table 3.2b

#### QLDACCM : Predicted values & 95% confidence intervals for 1991 values using model in Table 3.2a

month	predicted	confidenc <del>e</del>	limits	actual	error
January	0.69	0.19	1.20	20/31 = 0.65	- 0.04
February	0.80	0.28	1.32	19/28 = 0.68	- 0.12
March	0.89	0.35	1.43	34/31 = 1.10	0.21
April	0.76	0.20	1.32	28/30 = 0.93	0.17
May	0.81	0.24	1.38	37/31 = 1.19	0.38
June	0.92	0.33	1.51	25/30 = 0.83	- 0.09
July	0.91	0.30	1.51		
August	1.02	0.40	1.64		
September	1.08	0.45	1.72		
October	0.97	0.32	1.62		
November	1.04	0.38	1.71		
December	0.98	0.30	1.66		

#### Table 3.3a

#### SAACCM : Estimates for 'airline' model

	Jan 1976 - Dec 1990		
	estimate	std error	t value
MA (1)	0.67	0.059	11.5
SMA (12)	0.63	0.063	10,1
Noise std devn	0.25		
Residuals			
skewness	0.39		2.07
kurtosis	0.61		1.61

chi squared statistic = 19.3 (p = 0.37)

#### Table 3.3b

## SAACCM : Predicted values & 95% confidence intervals for 1991 using model in Table 3.3a

month	predicted	confidence	limits	actual	error
January	0.41	0	0.83	10/31 = 0.32	- 0.09
February	0.37	0	0.82	17/28 = 0.61	0.24
March	0.38	0	0.85	19/31 = 0.61	0.23
April	0.39	0	0.88	15/30 = 0.50	0.11
May	0.31	0	0.81	11/31 = 0.35	0.04
June	0.33	0	0.85	9/30 = 0.30	- 0.03
July	0.27	0	0.81		
August	0.37	0	0.93		
September	0.48	0	1.05		
October	0.49	0	1.07		
November	0.39	0	0.99		
December	0.53	0	1.15		

----

#### Table 3.4a

	Jan 1976 - Dec 1990 estimate stderror t valu		
	0.00	0.000	
MA (I)	0.92	0.033	28.2
SMA (12)	0.71	0.055	13.0
Noise std dev	0.10		
Residuals			
skewness	-0.06		-0.32
kurtosis	1.85		4.87
chi squared statistic = $10.9 (p = 0.89)$			

#### TASACCM : Estimates for 'airline' model

#### Table 3.4b

#### TASACCM : Predicted values & 95% confidence intervals for 1991 using the model in Table 3.4a

month	predicted	confidence	limits	actual	error
January	0.20	0	0.41	4/31 = 0.13	- 0.07
February	0.17	0	0.38	5/28 = 0.18	0.01
March	0.18	0	0.40	8/31 = 0.26	0.08
April	0.16	0	0.37	6/30 = 0.20	0.04
May	0.12	0	0.32	7/31 = 0.23	0.09
June	0.16	0	0.37	7/30 = 0.23	0.07
July	0.11	0	0.32		
August	0.15	0	0.36		
September	0.16	0	0.37		
October	0.13	0	0.33		
November	0.21	0	0.41		
December	0.19	0	0.40		

#### Table 3.5a

#### VICACCM : Estimates for 'airline' model

	<b>Jan 1976 - Dec 1990</b>		
	estimate	std error	t value
MA (1)	0.80	0.047	17.0
SMA (12)	0.70	0.058	12.0
Noise std dev	0.29		
Residuals			
skewness	0.05		0.27
kurtosis chi squared statistic = $16.5 (p = 0.56)$	0.14		-0.38

#### Table 3.5b

#### VICACCM : Predicted values & 95% confidence intervals for 1991 using the Model in Table 3.5a

month	predicted	confidence	limits	actual	error
January	1.20	0.63	1.77	40/31 = 1.29	0.09
February	1.26	0.69	1.84	37/28 = 1.32	0.06
March	1.52	0.93	2.11	54/31 = 1.74	0.22
April	1.19	0.58	1.79	25/30 = 0.83	- 0.36
May	1.45	0.84	2.07	31/31 = 1.00	- 0.45
June	1.36	0.73	1.98	26/30 = 0.87	- 0.49
July	1.07	0.44	1.71		
August	1.05	0.41	1.70		
September	1.33	0.68	1.98		
October	0.99	0.32	1.65		
November	0.99	0.32	1.69		
December	1.29	0.60	1.97		

#### Table 3.6a

#### WAACCM : Estimates for 'airline' model

	Jan 1976 - Dec 1990		
	estimate	std error	t value
MA (1)	0.86	0.038	22.5
SMA (12)	0.77	0.053	14.5
Noise std dev	0.17		
Residuals			
skewness	0.28		1.47
kurtosis	0.061		0.16
chi squared statistic = $27.9 (p = 0.06)$			

#### Table 3.6b

## WAACCM : Predicted values & 95% confidence intervals for 1991 using the Model in Table 3.6a.

month	predicted	confidence	limits	actual	error
January	0.49	0.16	0.82	23/31 = 0.74	0.25
February	0.50	0.17	0.83	10/28 = 0.36	- 0.14
March	0.51	0.17	0.84	12/31 = 0.39	- 0.12
April	0.51	0.17	0.85	17/30 = 0.57	0.06
May	0.44	0.09	0.78	11/31 = 0.35	- 0.09
June	0.54	0.19	0.88	11/30 = 0.37	- 0.17
July	0.50	0.15	0.85		
August	0.48	0.13	0.83		
September	0.57	0.21	0.92		
October	0.38	0.02	0.73		
November	0.52	0.16	0.89		
December	0.60	0.24	0.97		

\_ \_

.

#### 4. COMPARISON OF STATES AND AUSTRALIA

#### 4.1 ARIMA Models Fitted to Data up to December 1990

The 'airline' model appears to fit all states well and parameter estimates are close to the estimates for the national series; see Table 4.1 for a summary of estimates. However, individual states have substantially different patterns with

- NSW having an abrupt change in level occurring in 1983;
- Victoria having a change in level occurring at the end of 1979 and beginning of 1980;
- WA shows increases when NSW and Victoria are showing decreases (1979).

Thus in terms of **explanation**, the pure time series model has no merit since the same model fits all states well. The series ARIMA for the six states could be interpreted as being realisations of the same stochastic process, showing a degree of independence which **needs to be investigated in further work**. From Table 4.1 it is seen there is a similarity of estimates for states. A formal significance test shows that South Australia has an estimate MA (1) different from a common value.

#### 4.2 Aggregated Prediction Equations

In general, predictions for the ARIMA model are of the form

$$\hat{\mathbf{y}}_{t+k} = \sum_{j=1}^{n} a_{j,k} \mathbf{y}_{t-j}$$

where  $\hat{y}_{t+k}$  is the prediction of  $y_{t+k}$  given  $y_t, y_{t-1}, ..., y_n$  and the  $a_{j,k}$  depend on the values of the MA (1) and SMA (12) estimates. Since there is no reason to consider the MA (1) and SMA (12) estimates to be different for different states (except, perhaps, South Australia), the predictions for each state will be the same linear combinations of the lagged values for that state. These will, of course, be different, because the series are different from state to state. That is

$$\hat{y}_{t+k}^{S} = \sum_{j=1}^{n} a_{j,k} y_{t-j}^{S}$$

$$\hat{y}_{t+k}^{A} = \sum_{j=1}^{n} a_{j,k} y_{t-j}^{A}$$

$$= \sum_{j=1}^{n} a_{j,k} \left( \sum_{all \ S} y_{t-j}^{S} \right)$$

$$= \sum_{all \ S} \hat{y}_{t+k}^{S}$$

= aggregated predictions from individual states,

where superscript A refers to predictions and values for Australia and superscript S to those for the states. Thus the airline model with either

[1976 - 1990 data] MA (1) = 0.80, SMA (12) = 0.73or [1983 - 1990 data] MA (1) = 0.74, SMA (12) = 0.68

can be used to estimate future values for each state individually and the aggregated predictions are the same as those for the Australia series. This has not been investigated but could be the basis for futher work.

#### 4.3 Models for Jurisdictions with Small Counts

Time series models for series having small values, such as the series for Tasmania, Northern Territory and ACT, should be developed. The type of problem that needs to be addressed includes making proper allowance for the positive semi-definiteness of the observed values (so obtaining confidence intervals, for example, which are bounded below by 0), and the inappropriateness of a normal distribution approximation for the distribution of a small value.

#### 4.4 How Good are Predictions for 1991 Accident Data?

The prediction errors for the ARIMA models fitted to the Australian and state data for the period January to June 1991 are given in Table 4.3. The ARIMA models fitted to each state are very similar as we noted above. What we note is that for the states, the pattern of errors is more or less random for the months January to March. All errors for June, except Tasmania which has a small accident occurrence, are negative, that is the prediction is an overestimate, suggesting some structural change in the mechanism generating the data. Predictions for Victoria for April to June 1991 are all too large, indicating a structural change taking place.

To obtain some external assessment of how good the predictions are and how small the predictions errors, we can consider the NFRC in a Poisson model presented in Standardisation, section 1.2. If y is the number of accidents occurring in a month, the daily rate is  $\lambda$  and y follows a Poisson distribution, then the daily observed rate (y/n), where n is the number of days in the month, has mean  $\lambda$  and variance ( $\lambda/n$ ). In Table 4.3, we give for each of the jurisdictions, the average daily fatal accident rate based on the first six months data of 1991, i.e. an estimate of  $\lambda$ . Based on this value of the average daily rate, we give the value of  $\sqrt{\lambda/30}$ , the standard deviation of (y/30), the daily rate for a 30 day month, assuming the mean of (y/30) is  $\lambda$ . We see that these values are of similar sizes to the average absolute errors of prediction based on the errors of prediction for the first six months of 1991. What this suggests is that the errors of prediction are consistent with accurately estimated means for the daily rates and that the size of the error reflects the intrinsic Poisson variability in the monthly counts. That is, the ARIMA prediction model is obtaining as much accuracy as one could expect given the Poisson-like variability.

#### Table 4.1

State	MA (1) parameter	SMA (12) parameter
New South Wales	0.84 (0.04)	0.74 (0.06)
Queensland	0.72 (0.05)	0.73 (0.06)
South Australia	0.67 (0.06)	0.63 (0.06)
Tasmania	0.92 (0.03)	0.71 (0.06)
Victoria	0.80 (0.05)	0.70 (0.06)
Western Australia	0.86 (0.04)	0.77 (0.05)
AUSTRALIA	0.804 (0.05)	0.734 (0.06)

#### Summary of model estimates for states for 'airline' model, Jan 1976 - Dec 1990

(Standard errors in parentheses).

#### Table 4.2

#### Summary of model estimates for NSW & Australia for 'airline' model, 1983 - Dec 1990

State	MA (1) parameter	SMA (12) parameter				
New South Wales	0.84 (0.06)	0.67 (0.10				
Australia	0.74 (0.075)	0.68 (0.098)				

(Standard errors in parentheses).

.

#### Table 4.3

## Summary of prediction errors (actual - prediction) for the first six months of 1991.

Errors are extracted from Tables 2.1d, 3.1d, 3.2b, 3.3b, 3.4b, 3.5b, 3.6d. Table also gives theoretical standard deviations based on Poisson model.

	AUS	NSW	QLD	S A	TAS	VIC	WA
January	0.47	0.41	- 0.04	- 0.09	- 0.07	0.09	0.25
February	0.09	- 0.19	- 0.12	0.24	0.01	0.06	- 0.14
March	- 0.18	- 0.71	0.21	0.23	0.08	0.22	- 0.12
April	0.44	0.18	0.17	0.11	0.04	- 0.36	0.06
May	0.06	0.05	0.38	0.04	0.09	- 0.45	- 0.09
June	- 0.98	- 0.33	- 0.19	- 0.03	0.07	- 0.49	- 0.17
average error	- 0.02	- 0.10	0.07	0.08	0.03	- 0.16	- 0.04
average absolute error	0.37	0.31	0.19	0.12	0.06	0.28	0.14
average accident per day	4.99	1.61	0.90	0.45	0.20	1.18	0.46
Poisson model standard deviation	0.41	0.24	0.17	0.12	0.08	0.20	0.12

- ----

### PART 2

### **Models Involving Explanatory Variables**

#### Overview

A two-stage approach is taken to the development of explanatory models for the number of fatal road crashes (NFRC). A data analytic technique using the Bayesian Analysis of Time Series (BATS) is initially used to find explanatory variables which have some explanatory power. Having found these variables, reduced structural models are then fitted to selected models based on the explanatory variables found from the first stage. Both monthly and quarterly series were considered for both Australia and the states. For Australia and monthly data, a model with current month's fuel sales and fuel sales lagged by a month gave statistically significant results. The model with new motor vehicle registrations gave marginally statistically significant results. For the larger states, NSW, Victoria and Queensland, a weather index, based on a weighted average of rain days per month, gave marginally statistically significant results and, in the case of Queensland, statistically significant results. For quarterly data, the total sales of automotive fuel gave a statistically significant result for both the Australia and Victoria series. Additionally, a model involving both fuel sales and the percentage change in petrol prices gave statistically significant results for both Australia and Victoria. For other states with quarterly data, no model involving an explanatory variable gave statistically significant results which improved upon a pure time series model. The fuel model for the quarterly Australian series was investigated further and prediction equations developed which additionally involved forecasts for fuel.

Quarterly Predictions for 1991 based on an explanatory model with fuel using quarterly data were no better than predictions based on a pure time series model for monthly data.

#### 5. GENERAL METHODOLOGY

#### 5.1 Introduction

The primary objective of this part of the project is to develop equations to explain and predict the numbers of fatal road crashes (NFRC). Currently there are two general methodologies for analysing time series data such as NFRC. The first is based on the standard regression model but the noise has a time series structure, that is

$$y_{t} = \sum_{j=1}^{k} \beta_{j} x_{jt} + n_{t}$$
(5.1)

where  $y_t$  is the dependent variable at time t,  $t = 1, ..., T, \beta_j$  is the regression coefficient for the j<sup>th</sup> explanatory variable which has value  $x_{jt}$  at time t, and  $n_t$  follows a time series model such as an ARIMA (Box and Jenkins, 1970). The second is based on a regression model which allows for a stochastic time-varying evolution of trend, seasonal and regression parameters, which are sometimes called 'structural models'; see, for example, Harvey and Durbin (1986) for a brief review of these models and comparison with ARIMA model. To summarise Harvey and Durbin, the statistician should seek to identify the main observable features of the phenomena under study and should then attempt to incorporate in his model an explicit allowance for each of these main features. Visual inspection of graphs of time series usually reveal trends and seasonals as important observable features of the data, and it seems desirable to model these features explicitly. By analogy with usage in econometrics this procedure is called structural modelling. In a structural model of an economic system each component or equation is intended to represent a specific feature or relationship in the system under study. Sometimes it is convenient to transform the structural model into a particular alternative form for specific purposes, such as forecasting, and this is called the reduced form of the model. In the time series case it is possible to transform a linear structural model into an ARIMA model and this may then be referred to as the reduced form of the structural model.

The structural model takes the form

$$\mathbf{y}_{t} = \boldsymbol{\mu}_{t} + \boldsymbol{\gamma}_{t} + \boldsymbol{\varepsilon}_{t} \tag{5.2}$$

where  $\mu_t$ ,  $\gamma_t$  and  $\varepsilon_t$  are the trend, seasonal and irregular components respectively. The terms  $\mu_t$  and  $\gamma_t$  are allowed to evolve stochastically with time. For example,  $\mu_t$  might be modelled by a linear trend  $\alpha_t + \delta_t t$  but where

$$\alpha_t - \alpha_{t-1}$$
 = independent white noise,  
 $\delta_t - \delta_{t-1}$  = independent white noise,

and the terms  $\varepsilon_t$  also follow an independent white noise series.

The term  $\gamma_t$  represents seasonal effects and can also be modelled to allow for time varying effects. Methodology follows Harvey and Durbin (1986, §2) and West and Harrison (1989, Chapter 8). Instead of, say, representing seasonal monthly effects by individual terms for each month (a fully specified model), monthly effects are represented by trigonometric terms, sinusoids with wave lengths 12 months, 6 months, 4 months, 3 months, 2.4 months and a cosine of wavelength 2 months. This approach gives an opportunity for parsimonious modelling in terms of the harmonics of the basic 12 month sinusoid.

Explanatory variables can be added to the model to give

$$y_{t} = \mu_{t} + \gamma_{t} + \sum_{j=1}^{k} \beta_{j} x_{jt} + \varepsilon_{t}.$$
 (5.3)

In econometric modelling, the xs would be exogenous variables. A further development of this model is to allow the regression coefficients  $\beta_j$  to evolve with time also.

We have the model

$$y_{t} = \mu_{t} + \gamma_{t} + \sum_{j=1}^{k} \beta_{jt} x_{jt} + \varepsilon_{t}$$
(5.4)

with

 $\beta_{jt} - \beta_{jt-1} =$  independent white noise

for j = 1, ..., k. The noise across j = 1, ..., k for given t would have some correlation structure.

Generally, the model with time varying  $\beta_{jt}$  can be used for exploratory purposes finding those explanatory variables for which the regression coefficient takes statistically and scientifically significant values. Models of this form can, of course, also be used for explanatory purposes. However, for predictive purposes one would not want to use a model with time varying  $\beta$ s but use one with time constant  $\beta$ s, as, obviously, in the first case, the regression relationship is stochastic and therefore increases, generally speaking, the variability of forecasts.

Our general method of investigating regression models is as follows.

- 1. For a given explanatory variable or set of variables, first fit a model with time varying trend, seasonal and regression coefficients. (This is done using the package BATS as explained below).
- 2. Inspect a plot of the estimate of the regression coefficient,  $\beta_{jt}$ , and 95% confidence limits against t and determine whether the estimated value is significantly different from zero and relatively constant. Calculate various diagnostic statistics to determine the adequacies of the model. (Part of the standard BATS output).
- 3. For models with significant explanatory variables, estimate time constant regression effects by using a reduced form structural model. (This is done using the Genstat program, Payne *et al*, 1987 as explained below).

In order to determine sets of explanatory variables which would be likely to give reasonable models, a forward selection procedure was used, investigating sets of explanatory variables which individually give significant results.

#### 5.2 Fitting Procedures and Diagnostics

The time varying regression parameter structural model (5.4) was fitted to data using the package BATS (Bayesian Analysis of Time Series, see West, Harrison and Pole, 1987). A feature of the fitting process is that various statistics are calculated 'on-line'. That is, for example, the mean parameter,  $\mu_t$ , is estimated using data  $y_1, ..., y_t$ , to the current time t although, of course, an estimate of  $\mu_t$  based on all the data  $y_1, ..., y_T$  can be found. Similar comments hold for other parameters and especially the regression parameters  $\beta_{jt}$ . Thus inspecting the values of estimated parameters for the entirety of the series is an important part of the analysis. The PC based package BATS provides a number of numerical summaries of the fitted models as well as various plots. These include estimates of time varying parameters (trend, seasonal and regression parameters) and forecast mean square errors and one-step-ahead predictions.

The analyses are based on a Bayesian paradigm so that posterior means and standard deviations are found for parameters. Prior distributions need to be specified and these can be chosen in a 'neutral' way so as to allow the data to 'speak for themselves'.

The BATS program does not allow for the regression parameter  $\beta$  to be time constant so that an alternative program was used in this case. The Genstat program (see Payne *et al*, 1987) can be used to fit reduced form structural models as follows. If the level parameter,  $\alpha_t$ , the slope parameter,  $\delta_t$ , and the seasonal parameter,  $\gamma_t$ , follow the evolutionary models outlined earlier:

 $\alpha_t - \alpha_{t-1}$  = independent white noise,  $\delta_t - \delta_{t-1}$  = independent white noise,  $\gamma_t - \gamma_{t-1}$  = independent white noise,

where s is the seasonal period (s = 4 for quarterly data, s = 12 for monthly data) then models involving time constant regression parameters  $\beta$  can be fitted using the 'recipe' outlined in Table 5.1. There are five basic models which are fitted which allow for combinations of fixed or random level, slope and seasonal parameters. By fixed it is meant that the parameter is time constant, that is, formally, the noise increment in the evolutionary model specification has zero variance. If the parameter is fixed then a term for it has to be included in the set of explanatory variables. The models allow for a fully specified set of (s - 1) seasonal parameters with no reduction in model generality. The Genstat code is given in Appendix C.

#### 5.3 Diagnostics

As a check on model adequacy various diagnostics can be carried out. For the models fitted using the BATS package, graphical checks were made to investigate the constancy of regression parameter estimates with time. Other checks include patterns of residuals and autocorrelations. Residuals for these models can be defined in terms of one step-ahead prediction errors. Thus as a measure of the goodness-of-fit of the model, the estimated variance of the one step-ahead prediction error can be used, which we will

denote by  $S_{pe}^2$ . For the models fitted by BATS, the parameters are fitted using data up to and including the current value (i.e. 'on-line'), so that one step-ahead predictions are truly based on historical data, and are in themselves an independent check on the adequacy of the model. There is no need to check the models on independent data because future (relative to the current) values have not been used in the fitting process.

Other fitting procedures, such as ARIMA, use all the data to fit the parameters and then use this fitted model to predict 'future' values retrospectively from a 'current' value and past data.

The diagnostic statistics we provide include the following

(i)  $S_{pe}^2$ , the one step-ahead prediction error.

(ii)  $R_m^2$  calculated as

$$1 - \frac{n_v S_{pe}^2}{\sum(y_i - \overline{y})^2}$$

where  $n_v$  is the number of degrees of freedom on which  $S_{pe}^2$  is based. Note that because  $S_{pe}^2$  is calculated 'on-line', that is, using past data to predict a future value, it is quite possible for

$$n_v S_{pe}^2 > \sum (y_j - \overline{y})^2.$$

This would occur if there were a spurious relationship between dependent and explanatory variables and the overall mean were a reasonable predictor of values.

Thus this  $R_m^2$  can be negative! It would indicate a poorly specified model.

(Note the usual  $R^2$  is calculated as

$$\frac{1 - \frac{\text{sum of squares of residuals}}{\sum (y_i - \overline{y})^2}$$

and is necessarily non-negative).

(iii)  $R_{s}^{2}$  calculated as

1 -  $(n_v S_{pe}^2 / sum of squares of first differences around the seasonal means of first differences).$ 

This is a modification of a statistic introduced by Harvey (1990, Chapter 5.5.5).

This statistic refers  $S_{pe}^2$  to a sum of squares which has made allowance for a changing stochastic level (first differences) and fixed seasonal effect (seasonal means) and thus is a measure of the explanatory power of the model over and above that given purely by a trend + seasonal components time series model.

Thus the value of  $R_S^2$  is particularly important for our study. The larger the value of  $R_S^2$  the better. Large positive values of  $R_S^2$  indicate models which are considerably better than pure time series models. The value of  $R_S^2$  is zero if the model fitted is  $\nabla y_t$  = seasonal mean + independent error.

Again it is possible for  $R_S^2 < 0$  and for  $R_S^2 > R_m^2$ . Models for which  $R_S^2 < 0$  should be discarded because they give worse explanation than a simple time series model.

#### 5.4 Choice of Explanatory Variables

Traffic accident studies are obviously multidisciplinary and can benefit from the best of engineering, social and statistical sciences or suffer from the worst. One aspect of social science studies is that of 'data mining' or data dredging' to find statistically significant explanatory models. Given a sufficiently rich set of explanatory variables, serendipity should eventually throw up a significant explanatory effect where in reality there is none. In this study we have included explanatory variables which, *a priori*, were believed to have some effect on accident rates. Relationships found from data then have to have characteristics consistent with theory, whether it is economic, psychological or engineering, that is, for example, coefficients have to have the correct sign. Thus our modelling process has been to let the 'data speak for themselves', allowing for very flexible models in the first step. These models have, when they have been found to provide statistically and scientifically significant results, been simplified to provide parsimonious models by taking time dependent regression coefficients to be time constant.

#### 5.5 Transformations of Data

On empirical and statistical grounds, rather than perhaps theoretical road traffic grounds, many studies have transformed the dependent variables, NFRC and NCF (number of crash fatalities), using logarithms or square roots. From a statistical point-of-view, a first approximation to the distribution of NFRC would be Poisson (see Vol. 1, §8) leading to the relationship (mean = variance) for the distribution and a square root transformation of the dependent variable would be used to obtain a constant variance. See also Vo. 1, §4.5 for a discussion of 'Poisson regression' and associated difficulties. The other transformation frequently used in the literature is the logarithmic transformation. Here the basic idea generally is to obtain a multiplicative model for the mean of the original dependent variable because standard linear models are fitted to the logged dependent variable.

The recent statistical methodology Generalised Linear Models (see McCullagh and Nelder, 1989) allows for the separation of the relationship between mean and variance of the response (as in the case of Poisson data) and the relationship between the mean of the response and the linear predictor, that is  $\beta_1 x_1 + ... + \beta_p x_p$ . Advanced software such as BATS also allows for Stochastic evolution of parameters in Generalised Linear Models (see, also, West & Harrison, §10.6.4, 1989).

In this study, we feel that the signal to noise ratio is sufficiently small (we deal with NFRC on a monthly basis for states at the most disaggregated level) that from the data analysis point-of-view there is little evidence to prefer multiplicative models to additive models, that is

to

$$E(y) = exp(\beta_1 x_1 + ... + \beta_p x_p)$$

$$E(y) = \beta_1 x_1 + ... + \beta_p x_p.$$

Also because the relative ranges (max/min) of the values in the series over the periods of study do not generally exceed two, there is little to be lost by using statistical methods which assume that the variance of the response is constant, independent of the mean. For these reasons we have used additive models fitted by standard normal theory techniques to the original (or rescaled) responses throughout this study. Sometimes we have used logarithmic transformations to confirm our views put above about the insensitivity of the analysis to the choice of additive or multiplicative model.

#### 5.6 Degree of Aggregation for Explanatory Variables

Values of series are available at different levels of aggregation. For example, the period of the series can be monthly, quarterly, annual or longer. The spatial aggregation of a series can be regional, state or national. For this study, monthly and quarterly time periods were used and state and national spatial aggregations used. We have analysed series using the highest level of aggregation. We have not taken, say, a series which is collected at three yearly intervals and interpolated values to obtain an annual series. Our models all include trend terms and when these are stochastic, the interpolated values would be aliased to a large extent with the stochastic trend. Thus there would be little explanatory power to be derived from such a variable.

#### 5.7 Variables used in the Study

The dependent variables for the study, the number of fatal accidents and the number of fatal road traffic accidents in each State and territory, are available from July 1976 until June 1991 in the first case, from January 1970 until December 1990 in the second case. Details of the sources are given in the Appendix B.

For monthly analyses we consider the dependent variable to be the monthly count of fatal crashes divided by the number of days in the month, giving the same dependent variable as used earlier. For quarterly data we have standardised by population estimates so as to facilitate comparisons between states (a cross sectional approach). Where significant regression effects have been found we have carried out analyses using the raw quarterly count. Dividing by population tends to increase the downward trend of the series. Dividing by population is mentioned in §2.4, §4.2, §4.3 (cross sectional studies), §6.9 (through routes per thousand inhabitants) of Volume 1. In Section 9 of Volume 1, a summary table is given. Fatality rate per head of population is considered in the Legislation (references [32], [58]), Driver demographics (reference [58]), Vehicle demographics (references [27], [58]), Exposure (reference [45]), Weather (reference [58]).

As a summary, Section 9 of Volume 1 gives a list of explanatory variables and corresponding response variables. The first group is Legislation. When considering individual states then effects of legislation can be included as explanatory variables. However, the effects of legislation can be 'smeared' in various ways. Publicity about the political debate over legislation can be effective both on a national basis and prior to the time of actual enforcement of the legislation. So effects can be smeared across time

and space. For this reason, in our analyses of both national and state data we have not included variables in the Legislation group.

The second group is Economic factors. Of this group some are appropriate for single time cross sectional studies rather than longitudinal studies, and some for annual rather than monthly or quarterly data. We decided to obtain data on the following

fuel prices, consumer price index cost of transport (proxy for cost of an accident) (income, variation too small, omitted) gross domestic product unemployment [hospital access, variation too small, omitted].

#### The third group is Driver demographics

[strike days, effect in London study only, omitted] [holiday, confounded with seasonal effects, omitted, although Easter effect was tried] alcohol consumption, investigated - see below [young drivers; information on licensed drivers not readily available] [seatbelt usage, no reliable data available].

#### The fourth group is Vehicle demographics

[urbanisation, variation too small] [safety equipment, data not readily available] [traffic volume index, no reliable data] [speed, no reliable data] [vehicle mix, not investigated] **motor vehicle registrations** [single vehicle accidents, not used] [effect of RBT, not investigated] [highway capacity, not used].

The fifth group is Exposure

number of vehicles (new vehicle registrations used) population [VKT, no reliable data] car registrations (investigated) [length roads, not used] [routes, not used] [publicity, not used].

The sixth group is Weather

[temperature, not used] rain (weather index, weighted rain days) [snow, not used] [fog, not used] [wet road, not used] [wet weather index, see rain] [icy roads, not used] [nightime, not used].

Variables were not used because either they were thought inappropriate for this study, or no reliable data source is available.

Two additional variables were used to those given above. These are sales of fuel for motor vehicles and the relative change in petrol price. The first variable can be seen as both a proxy for VKT and an economic indicator. The second variable was considered because of the increases and decreases in petrol price that have taken place over the last decade. It might be thought that some discretional travel might be affected by short term changes in petrol prices. As far as we know changes in petrol price have not been used in explanatory models before except in first differencing of both response and explanatory variables i.e.

$$\nabla y = \nabla x + error.$$

Data for explanatory variables were obtained from various sources. Extra considerations in addition to those given above to bear in mind when assessing the possibility of using variables include length and continuity of available series; obtainability; frequency of measurements; jurisdiction for collection. The economic variables are

- gross domestic product (GDP);
- unemployment rate (UNEMP);
- consmer price index (CPI);
- consumer price index for transportation group (CPITRS);
- retail price of petrol (PETROL);
- real price of petrol, that is PETROL/CPI;
- change in price of petrol from one month to the next, calculated for month t as

$$\frac{\text{PETROL}_{t} - \text{PETROL}_{t-1}}{\text{PETROL}_{t-1}} \times 100$$
(% CHGPET).

The next group involves vehicle demographics.

- new motor vehicle registrations (MVR).

The next group is exposure involving a proxy for vehicle kilometres travelled (VKT).

- sales of automotive gasoline by state marketing area (FUEL);
- sales of inland automotive diesel oil by state marketing area (DIESEL);
- sales of LPG for automotive use by state marketing area (LPG).

The next group involves weather

- the weighted average by population of the number of rain days per month found at various locations within each state or territory giving a weather index for each state (RDAYS and WI).

Information on other variables which were investigated but not used is given in the Appendix B. One important omission from the above list of variables is the level of alcohol consumption since alcohol is claimed as a cause in many accidents. Alcohol consumption seems constant and inelastic to economic conditions and therefore offers no power as an explanatory variable in this study; see additional comments in Appendix B.

The categories of variables considered here, economic, vehicle stock, weather and vehicle usage, cover the main types of factors found in the literature which are appropriate to the study here.

# Table 5.1ARIMA model specifications for fitting various time constantregression models in Genstat

Model	level	slope	seasonal Y	<b>ARIMA</b> parameters						
	α	δ		Р	d	q	Р	D	Q	S
1	fixed	fixed	fixed	0	0	1				
2	random	fixed	fixed	0	1	1				
3	random	random	fixed	0	2	2				
4 (monthly)	random	random	random	0	2	2	0	1	1	12
4 (quarterly)	random	random	random	0	2	2	0	1	1	4
5 (monthly)	random	•	random	0	1	1	0	1	1	12
5 (quarterly)	random	-	random	0	1	1	0	1	1	4

#### 6. **RESULTS FOR MONTHLY SERIES**

#### 6.1 Australian Monthly Series

The dependent variable was taken to be the daily average fatal accident rate i.e. monthly NFRC divided by the number of days in the month giving AUSACCM in units of fatal accidents per day; four explanatory variables were investigated, UNEMP, MVR, FUEL and DIESEL. Additionally, variables lagged by one month were investigated. Values and plots of AUSACCM are given in Appendix A (Table A.1, Figure A.1). Series for all variables were available from March 1981 until December 1990, giving n = 105 data points.

We consider periods for fitting **explanatory** models which include early 1983 although this includes an abrupt change in level. Here we are trying to develop **explanatory** models using independent variables so this abrupt change provides a good opportunity for an explanatory variable to show whether it has explanatory power or not. To some extent, the AUSACCM series shows similar abrupt changes in level at the beginning of 1983 and at the beginning of 1990.

The preliminary data analysis using BATS was carried out using UNEMP, MVR, FUEL, DIESEL and lagger versions of these variables as explanatory terms. The only variable giving reasonable results was MVR. Various other combinations of the explanatory variables were investigated but no other models reduced MSE to any value

close to that for MVR in the two periods considered. Also values of  $R_S^2$  were small or negative.

The value of  $R_S^2$  is also positive for the two periods considered for MVR. The lagged (by one month) value (LAGMVR) of MVR was also investigated but had no effect. Similarly the lagged value of UNEMP was also investigated.

On the basis of this evidence the MVR model appeared to be the best because the MSE
was smallest, the coefficient was the correct sign, i.e. positive, and the estimate of  $\beta_t$  was more or less constant over the period. A plot of MVR is given in Figure A.2, Appendix A.

The five reduced structural models for MVR were fitted to the same series. Results are given in Table 6.2a.

From Table 6.2a, for none of the reduced structural models is the estimated regression coefficient statistically significant. Structural model 4 gives the largest regression coefficient but not the smallest innovation variance.

Models involving FUEL were also investigated further. Given that FUEL is a proxy for VKT there is obviously some delay in fuel sales by the retailer to the vehicle user and again a delay by the user until the fuel is used. Thus the term LAG FUEL was included in the model so that both the current month's fuel sales and the previous month's are represented in the model. Although the BATS analysis was not promising, giving a large mean square error and unsteady coefficients for LAG FUEL and FUEL reduced structural models were fitted to the monthly data involving FUEL and LAG FUEL. The results in Table 6.2b show statistically significant regression effects for LAG FUEL (Z = 3.1) and FUEL (Z = 2.1) for the reduced structural model 2, which is the best of the five. Some doubt must be expressed about the consistancy of the regression coefficient for LAG FUEL over time. This point is investigated later in §8.5.

Although not suggested as an important variable in the literature, some suggestion has been made in Australia (Queensland Transport Report, 14 May 1992) of the supposed importance of the Consumer Sentiment Index (CSI) (IAESR, University of Melbourne and Westpac). The CSI is compiled monthly by IAESR from a survey of 1200 people who are asked 5 questions relating to sentiment. The index is an average of the ratios of the number of favourable to unfavourable replies. The reduced structural model 5 was fitted with CSI as an explanatory variable and CSI found not to be significant.

#### 6.2 NSW Monthly Series

For NSW, the dependent variable, NSWACCM is the value of NFRC divided by the number of days in the month. For NSW following explanatory variables were considered: UNEMP, MVR, WI, FUEL, DIESEL. From the BATS analysis three models were found to be promising, see Table 6.3.

For these results in Table 6.3, it is seen that the model with WI looks reasonable giving

a large value of  $R_{S}^{2}$ , the correct sign for the estimate and the BATS analysis gives an estimate approximately constant over time. The model with DIESEL also appears

promising but the BATS analysis gives coefficients which show variability with time.

The five reduced structural models were fitted with WI as the explanatory variable. Results are given in Table 6.4. Plots of NSWACCM and WI are given in Appendix A, Figures A.3 and A.4.

Model 1 is a particularly simple model with fixed monthly effects, a fixed linear trend and MA(1) errors, and should be discounted. The other structural models all give estimated regression coefficients of the correct sign and similar size; WI being a weighted average of 'rain days' and thus having an expected negative effect on NFC. The regression estimate for model 3, random intercept and random trend, is most significant and has the equal smallest innovation variance.

# 6.3 Victorian Monthly Series

As for NSW, the Victoria state standardised series, VICACCM, was taken as the dependent variable. For Victoria the following explanations variables were considered: UNEMP, MVR, WI, FUEL, DIESEL. Promising models found from the BATS analyses are given in Table 6.5.

Consequent upon these results, the five reduced structural models were fitted for MVR and WI individually.

For the MVR model, Table 6.6, the estimated coefficients for models 2 to 5 are of the 'correct' sign, that is positive, but no regression coefficient is statistically significant. In terms of minimising the innovation variance, model 2 is the best. For the WI model with results given in Table 6.7, the best models appear to be models 2 and 3. Both have the 'correct' sign, negative, but the coefficients are not statistically significant. Plots of VICACCM, MVR and WI are given in Appendix A, Figures A.5, A.6 and A.7.

#### 6.4 Queensland Monthly Series

As for NSW and Victoria, the state standardised series, QLDACCM, was taken as the dependent variable. In Table 6.8 we give details of BATS models which were investigated.

The model given by WI looks promising because the value of  $R_S^2$  is reasonably large taking a value 0.15 in the first period and 0.18 in the second period studied. Recall that this is a measure of variability which is explained after allowance has been made for

both trend and seasonality. Other models are not interesting because  $R_S^2$  is close to 0 or negative.

In Table 6.9, we give the results of fitting the five reduced structured models with WI as the explanatory variable. It is seen that the regression coefficient is almost statistically significant (z value = -1.8) and negative, indicating that the fatal accident rates decrease with the increase in number of rain days. This appears to be a plausible relationship in Queensland where rainfall tends to be heavy. The smallest innovation variance is for model 1, but all innovation variances are similar for the five models as are the estimated regression coefficients.

This suggests that the simple model with fixed trend and seasonal effects is adequate to explain the variation in the QLDACCM series. A plot of QLDACCM and WI is given in Appendix A, Figures A.8 and A.9.

The variable CSI (see §6.1) was additionally used as an explanatory variable for the Queensland series and found not to be significant. This is not surprising as CSI is an Australia wide series and was not significant for the Australia series. However, Queensland Transport (Report, 14 May 1992) suggests CSI is an important variable, a finding not substantiated here.

# 6.5 Conclusions

For Australia, the best reduced structural model is given by FUEL + LAG FUEL with both regression coefficients being statistically significant and of the correct sign. The next best model is given by MVR and the regression coefficient for MVR, new motor vehicle registrations, by itself is found to be the correct sign but statistically not significant. Typical values of monthly FUEL sales in 1990 are of the order of 1.45 million kilo litres. The regression coefficient for FUEL is  $2.16 \times 10^{-6}$  and that for LAG FUEL is  $3.16 \times 10^{-6}$ . Adding these two coefficients together gives  $5.32 \times 10^{-6}$ . A change in FUEL of the order of 10 percent of the typical value, that is 0.145 million kilolitres, produces a change in AUSACCM of the size  $5.32 \times 10^{-6} \times 0.145 \times 10^{6}$  or 0.77 fatal accidents per day. The average rate for the first six months of 1991 (see Table 4.3) is 4.99 fatal accidents per day. Now 0.77 is 15 percent of 4.99 so that a change of 10 percent of the average of FUEL produces a change of 15 percent of the average of AUSACCM.

For NSW, the regression coefficient for WI, the weather index based on rain days, is negative but statistically not significant. For Victoria, two promising models involve MVR, new motor vehicle registrations, and WI, the weather index. Neither individually nor together did these variables give statistically significant estimated regression coefficients. For Queensland, the regression coefficient for WI was found to be negative and almost statistically significant. Thus, on a state basis, the variable WI has some but barely statistically significant explanatory power. It is interesting to compare the estimated regression coefficents for the three states NSW, Victoria and Queensland for the WI reduced structural models. They are given respectively by -11.1 (8.8), - 2.7 (6.5), - 10.7 (5.7) (values multiplied by  $10^3$  and standard errors in parentheses). We note a similarity in the estimated coefficients for NSW and Queensland taking values close to -  $11 \times 10^{-3}$ . This value corresponds to a reduction of 0.011 fatal accidents per day for each additional rain day. Given that the average number of accidents per day for NSW and Queensland are 1.61 and 0.90 for the first six months of 1991 (see Table 4.3) rain could have a practically significant effect reducing accidents by as much as about 30 per cent for Oueensland (0.9) to 0.6 for a 30 day rain month) and about 20 percent for NSW (1.6 to 1.3 for a 30 day rain month). On a national basis, WI has no real meaning (a weighted average of rain across major cities and towns) and was not computed for Australia.

#### Table 6.1

#### Various BATS models which were fitted to AUSACCM Period July 1979 - December 1990

X-variable(s)

		MSE	R <sup>2</sup>	$R_{s}^{2}$
UNEMP		0.83	< 0	< 0
MVR		0.71	0.11	0.10
UNEMP	MVR	1.37	< 0	< 0
UNEMP	LAGUNEMP	1.30	< 0	< 0
MVR LAG MVR		1.01	< 0	< 0
LAGUNEMP		1.09	< 0	< 0
	Period March 1981	l - December 1990		
MVR		0 54	< 0	0 17

MVR		0.54	< 0	0.17
FUEL		0.56	< 0	< 0
DIESEL		0.55	< 0	< 0
MVR	FUEL	0.64	< 0	< 0
UNEMP	FUEL	0.62	< 0	< 0
UNEMP	MVR FUEL	0.75	< 0	< 0
FUEL	LAG FUEL	0.88	< 0	< 0

# Table 6.2a

# Reduced Structural models; Y = AUSACCM, X = MVR Period March 1981 - December 1990

Model	β̂x 10 <sup>6</sup>	se x 10 <sup>6</sup>	innovation variance
1	- 6.4	10.5	0.41
2	6.8	12.9	0.37
3	6.5	13.1	0.38
4	10.4	13.2	0.46
5	8.2	13.0	0.44

# Table 6.2b

# Reduced Structural models; Y = AUSACCM, X = FUEL + LAG FUEL

Model	Term	β̂x 10 <sup>6</sup>	se x 10 <sup>6</sup>	innovation variance
1	FUEL LAG FUEL	1.47 2.45	1.13 1.14	0.40
2	FUEL LAG FUEL	2.16 3.16	1.02 1.02	0.33
3	FUEL LAG FUEL	-	:	-
4	FUEL LAG FUEL	2.20 2.96	1.01 1.01	0.41
5.	FUEL LAG FUEL	2.25 3.02	1.01 1.00	0.39

1

# Table 6.3

# Details of various BATS models which were fitted to NSWACCM Period July 1979 to December 1990

# X-variable(s)

		MSE	<b>R</b> <sup>2</sup>	$R_s^2$
UNEMP		0.29	< 0	< 0
MVR		0.23	0.06	0.12
WI		0.18	0.29	0.27
MVR	WI	0.23	0.06	0.12
MVR	LAGMVR	0.27	< 0	< 0
UNEMP	LAGUNEM	0.38	< 0	< 0

# Period March 1981 - December 1990

UNEMP	FUEL	0.22	< 0	< 0
MVR		0.21	< 0	< 0
WI		0.15	0.02	0.26
FUEL		0.19	< 0	0.09
DIESEL		0.19	< 0	0.23
WI		0.19	< 0	0.08
WI	FUEL MVR	0.23	< 0	0.10
FUEL	UNEMP	0.22	< 0	0.11
WI	UNEMP	0.23	< 0	0.08
WI	MVR	0.21	< 0	0.16

# Table 6.4

.

# Reduced Structural Models: Y = NSWACCM, X = WI Period March 1981 - December 1990

Model	$\hat{\beta} \times 10^3$	se x 10 <sup>3</sup>	innovation variance
1	13.1	9.4	0.14
2	- 10.4	8.7	0.14
3	- 11.1	8.8	0.14
4	- 9.3	8.6	0.17
5	- 9.0	8.6	0.16

# Table 6.5

#### Details of various BATS models which were fitted to VICACCM Period March 1981 to December 1990

# X-variable(s)

		MSE	<b>R</b> <sup>2</sup>	$R_s^2$
UNEMP		0.103	< 0	< Ŭ
MVR		0.079	< 0	0.07
WI		0.072	< 0	0.15
FUEL		0.103	< 0	< 0
DIESEL		0.102	< 0	< 0
WI	FUEL	0.107	< 0	< 0
WI	FUELMVR	0.114	< 0	< 0
FUEL	UNEMP	0.112	< 0	< 0
WI	UNEMP	0.108	< 0	< 0
WI	MVR	0.083	< 0	0.10

# Table 6.6

# Reduced Structural Models: Y = VICACCM Victoria, X = MVR Period March 1981 to December 1990

Model	$\hat{\beta} \times 10^5$	se x 10 <sup>5</sup>	innovation variance
1	- 0.28	1.17	0.063
2	0.39	1.22	0.060
3	0.41	1.24	0.061
4	0.47	1.24	0.073
5	0.44	1.21	0.070

# Table 6.7

# Reduced Structural Models: Y = VICACCM, X = WI Period March 1981 to December 1990

Model	$\hat{\beta} \times 10^3$	se x 10 <sup>3</sup>	innovation variance
1	- 0.19	6.8	0.063
2	- 2.64	6.5	0.060
3	- 2.77	6.5	0.061
4	- 1.25	6.5	0.073
5	- 1.40	6.5	0.070

## Table 6.8

# Details of various BATS models which were fitted to QLDACCM Period July 1979 to December 1990

#### X-variable(s)

		MSE	<b>R</b> <sup>2</sup>	$R_{s}^{2}$
UNEMP		0.075	< 0	< 0
MVR		0.067	< 0	< 0
WI		0.056	0.15	0.15
MVR	WI	0.066	0.00	0.01
MVR	LAGMVR	0.080	< 0	< 0
UNEMP	LAGUNEMP	0.105	< 0	< 0

# Period March 1981 to December 1990

# X-variable(s)

		MSE	<b>R</b> <sup>2</sup>	$R_s^2$
UNEMP		0.078	< 0	< 0
MVR		0.071	< 0	< 0
WI		0.056	0.11	0.18
FUEL		0.079	< 0	< 0
DIESEL		0.083	< 0	< 0
WI	FUEL	0.078	< 0	< 0
WI	FUELMVR	0.076	< 0	< 0
FUEL	UNEMP	0.080	< 0	< 0
WI	UNEMP	0.089	< 0	< 0
WI	MVR	0.067	< 0	0.03

# Table 6.9

# Reduced Structural Models: Y = QLDACCM, X = WI Period March 1981 to December 1990

Model	$\hat{\beta} \times 10^3$	se x 10 <sup>3</sup>	innovation variance
1	- 9.8	5.8	0.043
2	- 10.7	5.7	0.046
3	- 10.7	5.7	0.047
4	- 10.8	5.8	0.048
5	- 10.7	5.7	0.047

# 7. RESULTS FOR QUARTERLY SERIES

# 7.1 Introduction

For the quarterly series, the data were standardised by the number of days in the quarter and by an estimate of the population of jurisdiction, giving accident rates per day per 100,000 population. Generally, all explanatory variables used are standardised by population, e.g. unemployment rate (UNEMP), or are dimensionless such as the change in petrol price (% CHGPET). Standardisation by population also allows comparisons of estimated regression parameters amongst jurisdictions to be made. However, standardisation by population or not appears to have little effect on statistical significance of models.

Because some variables are only available on a quarterly basis, the accident data were aggregated by quarters and explanatory models fitted. An alternative approach is to interpolate quarterly series to obtain monthly values but this approach was not pursued because of the inherent problems of confounding interpolated values with stochastic trend parameters.

# 7.2 Australia Quarterly Series

For the Australia quarterly series, both monthly and quarterly series could be used as explanatory variables, aggregating or averaging as appropriate monthly series to obtain quarterly series. The variables used were %CHGPET, percentage change in petrol price, FUEL sales (in volume) of automative fuel, DIESEL sales (in volume) of automotive fuel, MVR, new motor vehicle registrations standardised by population, UNEMP, the unemployment rate. The variable %CHGPET is present to reflect short term economic and behavioural change to petrol price changes, FUEL and DIESEL are proxies for VKT, vehicle kilometers travelled, MVR is present to reflect short term economic changes as reflected by renewals of the vehicle stock, UNEMP is an indicator

of general economic activity. From Table 7.1 we see that some models have reasonably

large  $R_{S}^{2}$  values and these are FUEL, %CHGPET + FUEL, FUEL + MVR. These models were investigated further by fitting the reduced structural models and results are given in Table 7.2. The first model in Table 7.2 involves %CHGPET and gives no significant results. From Table 7.2 we see that statistically significant results are given by the FUEL model with Z values in excess of 3 for models 2, 4 and 5. Note that model 1, a simple regression model with fixed trend and seasonal effects with MA(1)errors, gives an insignificant regression estimate. The estimates and standard errors for the estimated regression parameter for FUEL are very similar across models 2, 4 and 5. For model 3 the estimation procedure did not converge and results are therefore not given. The estimated regression coefficient is positive indicating the correct sign. The next model involves FUEL and MVR, and the coefficient of FUEL is little changed for structural models 2, 4 and 5, from the model with FUEL by itself. This is reassuring. The estimated regression coefficient for MVR is not statistically significant for structural models 2, 4 and 5 (Z value equals about -0.4 to -0.5). Additionally the value of the regression coefficient for MVR is negative which is the wrong sign. (It could be argued that the coefficient for MVR should be negative if one associates MVR with the quality of the vehicle stock - new vehicles replacing old). This model involving FUEL and MVR therefore is ambiguous to interpret and is not recommended although statistically significant.

The last model involves FUEL and %CHGPET. The structural models 1 and 5 have the smallest innovation variances, however model 5 gives consistency with the model with FUEL by itself. Model 1 gives an almost zero and statistically insignificant regression coefficient for FUEL and almost significant regression coefficient for %CHGPET (Z = -1.6) with the correct sign, that is negative. Model 5 gives a statistically significant (Z = 3.1) regression coefficient for FUEL and statistically insignificant regression estimate for %CHGPET (Z = -0.40) which has the correct sign, negative.

Overall, the best models, in terms of minimising innovation variance and statistically significant and correctly signed regression coefficients are given below.

X = FUEL, structural model 2,

$$\hat{\beta} = 18.0 \times 10^{-9}$$
, se = 5.2 x 10<sup>-9</sup>, Z = 3.5

innovation variance =  $4.4 \times 10^6$ 

$$R_{s}^{2} = 0.526$$

X = FUEL + % CHGPET, structural model 5

FUEL

%CHGPET

$$\hat{\beta} = 18.0 \times 10^{-9}$$
, se = 5.8 x 10<sup>-9</sup>, Z = 3.10  
 $\hat{\beta} = -0.28 \times 10^{2}$ , se = 0.69 x 10<sup>2</sup>, Z = 0.40

innovation variance =  $4.8 \times 10^{-6}$ 

$$R_{\rm s}^2 = 0.363.$$

The FUEL only model is better than the FUEL + %CHGPET model in terms of innovation variance and significance of FUEL. Structural model 2, used with the best FUEL model, involves a random level and fixed seasonal (quarterly) effects.

In order to confirm that the model with FUEL in it gives a satisfactory fit, we give in Figures 7.1, 7.2 and 7.3 confirmatory plots and diagnostics from the BATS output. Figure 7.1 gives the plot of the time dependent regression coefficient estimate against time and it is seen to be very steady. Figure 7.2 gives the plot of residual standard deviation which steadily decreases and Figure 7.3 gives the sample autocorrelation function for residuals; these latter two plots give satisfactory diagnostics indicating an adequate model.

#### 7.3 Victoria Quarterly Series

For Victoria the results of the BATS analyses are given in Table 7.3. For no model was the estimated time dependent regression coefficient constant over time suggesting some lack of stability of relationship. Reduced structural models were then fitted to the better

models (with  $R_S^2 > 0.2$ ) of Table 7.3 and results for these estimated models are given in Table 7.4. In Table 7.4 we see that the model with FUEL as the explanatory variable gives statistically significant and correctly signed estimates for structural models 2 and 3 (Z = 2.4, 2.3, respectively). Structural model 2 appears best overall.

ANP 1991/FORS 27.03.92

The model with FUEL + %CHGPET provides constrasting estimates when grouping structural models 1, 4 and 5 and structural models 2 and 3 together. For models 1, 4 and 5, FUEL has statistically insignificant values ( $Z \sim 0$ ) of the regression parameter, and %CHGPET is not significant but with the correct sign (Z = -0.9, -0.4, -0.6) whereas for groups 2, 3, FUEL is statistically significant (Z = 2.2, 2.1) but %CHGPET statistically insignificant ( $Z \sim 0$ ). There is obviously a reasonably complex relationship between FUEL and %CHGPET but we have not investigated this further. In microeconomic terms, one would expect the variable %CHGPET to be causing changes in FUEL. In terms of minimising innovation variance and maximising significance of regression coefficients, models 2 and 3 are the best.

The model FUEL + MVR gives insignificant results for FUEL whereas MVR gives an insignificant (Z = 0.9) and correctly signed regression coefficient for structural model 3.

The model %CHGPET + DIESEL gives results similar to the model FUEL + %CHGPET except that structural models 2 and 3 give correctly signed but insignificant regression parameter estimates for both %CHGPET (Z = -0.7, -0.7) and DIESEL (Z = 1.2, 1.1).

Overall the best models are

X = FUEL structural model 2,  $\hat{\beta} = 7.8 \times 10^{-8}$ , se = 3.3 x 10<sup>-8</sup> Z = 2.4 innovation variance = 1.21 x 10<sup>-5</sup>  $R_s^2 = 0.639$ 

X = FUEL + %CHGPET structural model 2,

FUEL  $\hat{\beta} = 7.8 \times 10^{-8}, \text{ se} = 3.5 \times 10^{-8}$  Z = 2.2%CHGPET  $\hat{\beta} = 3.0 \times 10^{-3}, \text{ se} = 8.5 \times 10^{-3}$  Z = -0.35innovation variance = 1.25 x 10^{-5}  $R_s^2 = 0.639$ .

# 7.4 NSW Quarterly Series

For NSW the results of the BATS analysis are given in Table 7.5. From this table we see that there are no models which are satisfactory as all values of  $R_S^2$  are negative. A

positive value of  $R_S^2$  would indicate some explanatory power of a variable after trend and seasonality has been allowed for. No reduced structural models were fitted and we conclude that there is no better explanatory model than a pure time series model.

# 7.5 Queensland Quarterly Series

For Queensland, the results of the BATS analyses are given in Table 7.6. All values of

 $R_{S}^{2}$  are negative, like the results for NSW. No satisfactory explanatory models could be found and our conclusions are similar to those for NSW.

#### Table 7.1

# Details of various BATS models which were fitted to AUSACCQ quarterly data Period 2nd quarter 1981 until 4th quarter 1990

#### X-variable(s)

		MSE x 10 <sup>6</sup>	R <sup>2</sup>	R <sup>2</sup> <sub>s</sub>
%CHGPET		11.7	0.32	< 0
FUEL		7.3	0.58	0.09
DIESEL		8.5	0.51	< 0
MVR		21.1	< 0	< 0
UNEMP		11.6	0.3	< 0
%CHGPET	FUEL	5.8	0.63	0.31
%CHGPET	DIESEL	8.5	0.46	< 0
FUEL	MVR	5.7	0.64	0.32
UNEMP	FUEL	8.4	0.48	0.01
%CHGPET	MVR	10.4	0.30	< 0
%CHGPET	FUEL MVR	21.2	< 0	< 0
FUEL	MVR LAGMVR	12.7	0.2	< 0
%CHGPET	FUEL UNEMP	7.0	0.59	0.08
%CHGPET	FUEL UNEMP MVR	15.3	< 0	< 0

	Australia Qu	arterly: Redu	iced st	ructural	models Y	= AU	JSACCQ
	X variables		1	2	3	4	5
1.	%CHGPET	$\hat{\boldsymbol{\beta}} \mathbf{x} 10^2$	- 1.30	- 0.98	- 0.96	- 1.02	- 0.99
		se x 10 <sup>2</sup> innovation	0.94	0.80	0.81	0.82	0.81
		variance x 10 <sup>6</sup>	6.5	6.3	6.5	8.8	7.1
2.	FUEL	$\hat{\beta} \times 10^9$	-	18.0	-	18.0	18.0
		se x 10 <sup>9</sup> innovation	-	5.2	-	5.3	5.2
		variance x 10 <sup>6</sup>	-	4.4	-	5.2	4.6
3.	FUEL	$\hat{\beta} \ge 10^9$	0.1	19.0	•	20.0	19.0
		se x 10 <sup>9</sup>	6.7	5.8	•	5.9	5.9
	MVR	β̂ x 10	0.41	- 2.48	-	- 3.37	- 2.53
		se x 10 innovation	6.4	6.2	-	6.7	6.3
		variance x 10 <sup>6</sup>	5.5	4.6	-	5.3	4.8
4.	FUEL	$\hat{\beta} \ge 10^9$	0.0023	1.40	-	18.0	18.0
		se x 10 <sup>9</sup>	6.6	6.7	-	5.8	5.8
	%CHGPET	$\hat{\beta} \times 10^2$	- 1.4	- 0.92	-	- 0.25	- 0.28
		se x 10 <sup>2</sup> innovation	0.9	0.79	-	0.72	0.69
		variance x 10 <sup>6</sup>	4.9	6.2	-	5.4	4.8

Note: where no results are given convergence of the fitting process did not occur.

# Details of various BATS models fitted to standardised quarterly data for Victoria, VICACCQ Period 2nd quarter 1981 until 4th quarter 1990

X-variable(s)

		MSE x 10 <sup>5</sup>	<b>R</b> <sup>2</sup>	$R_s^2$
%CHGPET		2.3	< 0	0.19
FUEL		2.1	< 0	0.28
DIESEL		2.3	< 0	0.21
MVR		2.6	< 0	0.10
UNEMP		3.1	< 0	< 0
%CHGPET	FUEL	2.1	< 0	0.31
%CHGPET	DIESEL	2.1	< 0	0.29
FUEL	MVR	2.1	< 0	0.30
DIESEL	MVR	2.2	< 0	0.28
%CHGPET	MVR	2.5	< 0	< 0

Victoria Quarterly: Reduced structural models $Y = VICA$	CCQ
--	-----

			Structural Model				
	X variables		1	2	3	4	5
1		$\hat{B} = 10^8$	0.0	7 9		0.0	0.0
1.	FUEL	p x 10	0.0	7.0	1.1	0.0	0.0
		innovation	3.6	3.3	3.4	3.9	3.9
		variance x 10 <sup>5</sup>	1.36	1.21	1.26	1.94	1.56
2	FILEI	$\hat{B} \times 10^8$	0.00	78	76	0.2	0.0
4.	TOLL	p x 10 so x 108	0.00 7 0	7.0	2.6	0.2	0.0
		se x 10°	3.8	3.5	3.0	4.1	4.0
	%CHGPET	$\hat{\beta} \times 10^3$	- 8.2	- 0.3	- 0.4	- 3.6	- 5.0
		se x 10 <sup>3</sup>	9.6	8.5	8.5	9.1	9.2
		innovation					
		variance x 10 <sup>5</sup>	1.37	1.25	1.32	2.02	1.62
2		â 10 <sup>8</sup>					0.0
3.	FUEL	108	0.0	0.2	0.1	0.0	0.0
		se x 10°	3.7	3.7	3.8	4.1	4.0
	MVR	β̂ x 10	- 3.0	3.3	5.8	6.2	3.7
		se x 10	5.8	6.7	6.7	7.0	6.9
		innovation					
		variance x 10 <sup>5</sup>	1.39	1.49	1.66	1.92	1.58
A	Ø CUCDET	$\hat{B} \times 10^3$	01	63	60	4.1	4.0
4.	%CHUPEI	μ x 10	- 8.1	- 0.5	- 0.0	- 4.1	- 4.9
		se x 10 <sup>-2</sup>	9.3	8.8	8.7	9.0	8.9
	DIESEL	$\hat{\beta} \ge 10^8$	- 0.1	6.2	5.6	0.0	0.0
-		se x 10 <sup>8</sup>	51	51	52	5.8	53
		innovation	J.1	J,1	<i>يە</i> . ئ	5.0	5.5
		variance x 10 <sup>5</sup>	1.37	1.41	1.48	2.06	1.63

# Details of BATS models fitted to standardised quarterly data for NSW, NSWACCQ Period 2nd quarter 1981 until 4th quarter 1990

# X-variable(s)

		MSE x 10 <sup>5</sup>	<b>R</b> <sup>2</sup>	$\mathbf{R}_{s}^{2}$
%CHGPET		3.05	< 0	< 0
FUEL		3.00	< 0	< 0
DIESEL		2.98	< 0	< 0
MVR		3.11	< 0	< 0
UNEMP		3.26	< 0	< 0
%CHGPET	FUEL	2.70	< 0	< 0
%CHGPET	DIESEL	2.21	0.16	< 0
%CHGPET	UNEMP	2.53	0.03	< 0
%CHGPET	MVR	3.82	< 0	< 0
%CHGPET	FUEL UNEMP	3.82	< 0	< 0
%CHGPET	DIESEL UNEMP	2.81	< 0	< 0

#### Table 7.6

# Details of BATS models fitted to standardised quarterly data for Queensland, QLDACCQ Period 2nd quarter 1981 until 4th quarter 1990

#### X-variable(s)

		MSE x 10 <sup>5</sup>	<b>R</b> <sup>2</sup>	$R_s^2$
%CHGPET		3.82	0.28	< 0
FUEL		7.35	< 0	< 0
DIESEL		11.8	< 0	< 0
MVR		9.48	< 0	< 0
UNEMP		8.82	< 0	< 0
%CHGPET	FUEL	6.69	< 0	< 0
FUEL	MVR	9.60	< 0	< 0
%CHGPET	UNEMP	7.43	< 0	< 0
FUEL	UNEMP	7.87	< 0	< 0
%CHGPET	MVR	4.42	< 0	< 0
%CHGPET	MVR FUEL	8.39	< 0	< 0



ON-LINE COEFFICIENT OF FUEL

Figure 7.1



ON-LINE S.D.

Figure 7.2



.

Figure 7.3

.

.

.

#### 8. EXPLANATORY & PREDICTIVE MODELS FOR ROUTINE USE

# 8.1 Introduction

In this section we critically review various models investigated earlier and provide models for prediction and explanation. In particular we give estimates of reductions in numbers of fatal crashes for 1990 and 1991 which are due to continuing road safety measures or economic conditions; see Section 8.8.

Since carrying out the preliminary work described in earlier sections, revised data for FUEL was obtained from ABARE and the most recent data available for other variables is used in analyses in this Section. Thus there are small differences in estimates when comparing models fitted in this section with those used in earlier sections. FUEL was available until Q3, 1991, monthly fatal crash figures until December 1991.

In Table 8.1 we summarise the best monthly and quarterly models which have been

found in the earlier sections. The two significant statistics are the 'Z-value' and the  $R_s^2$  value. The 'Z-value' indicates whether the estimated regression parameter is statistically significant from zero. We would expect a reasonable model to have a 'Z-value' in

excess of 2 in absolute size. The  $R_s^2$  value indicates how much better the explanatory model is compared with a pure time series model.

For monthly Australian data, the FUEL + LAG FUEL model has estimated regression

parameters which are statistically significantly different from zero. The value of  $R_s^2$ , 0.51, indicates a model with superior performance over a pure time series model. The model for Queensland is close to being significant. The values of the regression coefficients can be compared for Australia and the states. For MVR, the Australia and Victoria series give regression coefficients which are similar 6.8 x 10<sup>-6</sup> and 3.9 x 10<sup>-6</sup>, respectively. For WI, the regression coefficients for NSW and Queensland are almost identical, - 1.0 x 10<sup>-2</sup> and - 1.1 x 10<sup>-2</sup> which gives some support to the model although the monthly series are not standardised by population.

For quarterly data, the FUEL model for Australia gives a highly significant regression

parameter (Z-value = 3.5) and a value of  $R_s^2 = 0.526$  indicating superior performance over a pure time series model. Victoria gives similar results but the significance of the regression parameter is less and  $R_s^2$  is greater. The models with

FUEL and %CHGPET give little improvement in terms of  $R_S^2$  over the models with FUEL only and the regression coefficient for %CHGPET is not significantly different from zero. The regression coefficients for FUEL for the Australia and Victoria series are 1.8 x 10<sup>-8</sup> and 7.8 x 10<sup>-8</sup>, which are substantially different. The regression coefficient for %CHGPET, in the FUEL + %CHGPET model, is - 9.2 x 10<sup>-3</sup> for Australia and - 3.0 x 10<sup>-4</sup> for Victoria, here a larger value for the Australia series. These systematic differences should not be present if relationships were similar because of the standardisation by population used for the quarterly series.

ANP 1991/FORS 27.03.92

#### 8.2 FUEL model for Australian Quarterly Series

Results presented in the rest of this section use the revised FUEL Series and the most recent crash figures, unless otherwise indicated.

Of the models presented in §8.1, the model for the Australian Quarterly Series involving FUEL appears best for both explanatory and predictive purposes. In Tables 8.2 we give detailed results of the Reduced Structural model 2 fitted to data from 2nd Quarter

1981 to 4th Quarter 1989. The value of  $R_S^2$  is 0.54 and is satisfactorily large indicating a good improvement of the explanatory model over a pure time series model. Also given are results for the Reduced Structural Model 1.

The fitted Reduced Structural model 2 is

 $\nabla \text{ AUSACCQ}_{t} = -0.000818 \text{ x t} \\ (0.00023) \\ - 0.001212 \text{ x } \nabla(\text{indicator for quarter 2}) \\ (0.00091) \\ + 0.000315 \text{ x } \nabla(\text{indicator for quarter 3}) \\ (0.00091) \\ + 0.000340 \text{ x } \nabla(\text{indicator for quarter 4}) \\ (0.0013) \\ + 2.2 \text{ x } 10^{-8} \text{ x } \nabla \text{FUEL}_{t} \\ (0.56 \text{ x } 10^{-8}) \\ + \text{ MA error}$ 

with t being an index t = 1, 2, ... indicating the periods in the series starting at 2nd Quarter 1981 with t = 1.

In Table 8.3a we give results for the model fitted up to the 4th Quarter 1990. This model is

 $\nabla \text{AUSACCQ}_{t} = -0.000854$ (0.000228)  $-0.000868 \times \nabla (\text{indicator for quarter 2})$ (0.000833)  $+ 0.000641 \times \nabla (\text{indicator for quarter 3})$ (0.000854)  $+ 0.00083 \times \nabla (\text{indicator for quarter 4})$ (0.00124)  $+ 2.1 \times 10^{-8} \times \nabla \text{FUEL}_{t}$ (0.54 x 10<sup>-8</sup>) + MA error

which gives values not statistically different from the previous fit to data up to 1989.

Again the value of  $R_s^2$ , 0.46, is satisfactorily large.

From Table 8.3, the estimated regression coefficient for FUEL is  $2.1 \times 10^{-8}$ , the average value for FUEL is  $3.97 \times 10^6$  minimum and maximum values are  $3.61 \times 10^6$  and  $4.39 \times 10^6$ , that is average value of FUEL plus and minus about 10 percent of the average value of FUEL, which is  $0.1 \times 4.0 \times 10^6$ . A change of FUEL from one quarter to the next,  $\nabla$ FUEL<sub>t</sub>, of  $0.4 \times 10^6$  contributes a value of  $0.4 \times 10^6 \times 2.6 \times 10^{-8}$  (2.6 x  $10^{-8}$  is the estimate of the FUEL term from Table 8.3a) or  $1.04 \times 10^{-2}$  or 0.0104 to the explanatory term. Now the average value of AUSACCQ, from Table 8.3, is 0.0409 so that a change in FUEL equal to 10 percent of the average value of FUEL produces a change in the accident rate equal to  $0.0104/0.0409 \times 100$  or 25 percent of the average accident rate. Thus the relationship between the accident rate and FUEL has significant practical importance with a multiplying effect of 2.5 in terms of changes in average value of FUEL and accident rate.

Also in analysing the revised FUEL data, we found that structural model 1 gave a large

value of  $R_{S}^{2}$  (0.65 from Table 8.2a, 0.53 from Table 8.3a). Results for the model are included for interest but it tends to have worse predictions than structural model 2,

although, of course, using retrospective measures such as  $R_S^2$ , it appears a good model.

Table 8.3c gives results of these models fitted to data upto Q3 1991. Results are similar to data fitted up to Q4 1989 and up to Q4 1990, but the FUEL regression coefficient for Reduced Structural model 2 is estimated as  $1.8 \times 10^{-8}$ , cf  $2.2 \times 10^{-8}$  to Q4, 1989; 2.1 x  $10^{-8}$ , O4, 1990.

In Table 8.2b we give forecasts for 1990 based on the models fitted to data up to the end of 1989 and the actual values of FUEL for 1990. The forecasts of AUSACCQ for structural model 2 for 1990 have typically a relative error of 6 percent. Errors of forecasts are all negative. The confidence limits appear somewhat conservative. In Table 8.3b forecasts for 1991 are produced based on the model and data up to the end of 1990 for AUSACCQ. The errors for structural model 2 are 0 to 2 significant figures. The values of FUEL for each quarter of 1991 are taken to be equal to those for the corresponding quarter of 1990. This is a simple approach to providing estimates or forecasts of future values of FUEL to be used in the prediction equation for AUSACCQ. In the next section we discuss this point further and consider some time series models for FUEL in order to provide better predictions. However, first we consider the effect of standardisation by population on the NFC data values. Recall that standardisation was done to facilitate comparisons across states.

For the raw NFC quarterly data, the reduced structural models were fitted to the series from 2nd quarter 1981 to the 3rd quarter 1991. For this series, the reduced structural

models 1 and 2 were the best and results are given in Table 8.4a. For model 1, the  $R_s^2$  value is 0.54 indicating good explanation by FUEL in addition to that purely explained

by a time series model. For model 2,  $R_S^2$  is 0.48 and the estimated coefficient of FUEL is smaller and less significant than for model 1. We prefer model 2 because of its better predictive power. Using the model 2 FUEL estimate (2.8 x 10<sup>-4</sup>), a change in FUEL equal to 10 percent of the average value of FUEL (4 x 10<sup>5</sup>) gives a change in raw quarterly NFC equal to 112 fatal crashes which is about 18 percent of the average quarterly NFC (630 crashes). The equivalent change for the population standardised series, AUSACCQ, reported earlier is 25 percent, slightly larger than for the raw data series.

The Reduced Structural Model 2 uses first differencing so that the model fitted (Table 8.4a) to the series is

 $\nabla$  raw NFC<sub>t</sub> =  $\nabla$  FUEL<sub>t</sub> x 2.78 x 10<sup>-4</sup> (0.64 x 10<sup>-4</sup>) - 9.47 -  $\nabla$ Q2 x 12.4 +  $\nabla$ Q3 x 12.6 +  $\nabla$ Q4 x 21.2 (3.15) (11.5) (11.9) (14.8)

Forecasts for 1990 and 1991 for NFC are given in Tables 8.4c and 8.4d; the former for 1990 using NFC data up to 4th Quarter 1989 and FUEL up to 4th Quarter 1990, and the latter for 1991 using NFC data up to 4th Quarter 1990 and FUEL for 1991 estimated by the corresponding quarter of 1990. These tables give results for NFC corresponding to those for AUSACCQ in Table 8.2b and Table 8.3b respectively. Errors of forecasts for structural model 2 are generally smaller and are all negative and on average have a relative error of about 7 percent for 1990 forecasts (Table 8.4c) and 5 percent for 1991 forecasts (Table 8.4d). In Table 8.4e, we give forecasts for 1991 based on raw NFC up to Q4 1990 and FUEL up to Q3 1991. On this occasion errors are both positive and negative.

#### 8.3 Time Series Model for FUEL for use in forecasts

A brief literature search was carried out to investigate what work had been done to predict road transport fuel demand in Australia and overseas. Fuel demand is, of course, of great interest in its own right for economic planning. Donaldson, Gillan and Jones (1990) presented work describing models for predicting **annual** fuel demand and is not useful to our study here. Elsewhere, time series models have been fitted to overseas data.

Here we explore possible ARIMA time series models in order to predict FUEL for incorporation into the reduced structural model to predict either AUSACCQ or the raw quarterly NFC series. A number of models were fitted to FUEL using the STATGRAPHICS package and a summary of the better models is given in Table 8.5. All the models fit the series well and diagnostics for model inadequacy are all negative. Interestingly, FUEL only requires first differencing, and not seasonal differencing as well, for the better ARIMA models.

Forecasts are given in Table 8.6 for 1991 values based on the model fitted to data up to the 4th Quarter 1990 and values up to the same time. For Model 2 of Table 8.5, forecasts for the first two quarters of 1991 are both high with a relative error of about -3 percent.

For FUEL Model 3 of Table 8.5 forecasts for 1991 are given in Table 8.6. Forecasts are high but errors are smaller than for Model 2 and it is preferred for inclusion in the AUSACCQ prediction model.

#### 8.4 Combining the FUEL Explanatory Model with the Time Series Model for FUEL to give forecasts for Australian Quarterly series.

We can now combine the FUEL explanatory model for fatalities (Section 8.2) with the FUEL time series model (Section 8.3) to give a prediction model for Australian fatalities on a quarterly basis.

We consider the following two models for the population standardised series AUSACCQ and the quarterly NCF series.

#### AUSACCQ

AUSACCQ modelled by structural model 2 with

X = FUEL (see Table 8.3); FUEL modelled by a SARIMA (0, 1, 1, 1, 0, 0, 4) TSM (see Table 8.5).

#### Quarterly NFC

NFC modelled by structural model 2 with X = FUEL (see Table 8.4); FUEL modelled by a SARIMA (0, 1, 1, 1, 0, 0, 4) TSM (see Table 8.5).

Predictions for 1991 are given in Table 8.7. On the assessment of errors of prediction given in Table 8.4c, Table 8.4d and Table 8.7b for the raw NFC quarterly series, all errors are negative, that is predictions are too large. Table 8.4c looks at predictions for 1990 using actual FUEL for 1990, Table 8.4d looks at predictions for 1991 using FUEL from the corresponding quarter of 1990, Table 8.7b uses a TSM estimate of FUEL for 1991.

#### 8.5 Forecasts based on the monthly series

It is well known, (see, for example, Box and Jenkins, 1970, §5.1) quarterly forecasts can be derived from monthly forecasts by merely aggregating the monthly forecasts. Section 2 of this report considered time series models for the monthly series and some uniformity of results was obtained across analyses for Australia and the States. Below we give prediction errors (actual-prediction) for the quarterly forecasts for 1991 derived from Table 2.1d. The forecast for the 1st Quarter is given by

4.63 x 31 + 4.87 x 28 + 5.70 x 31 or 456.6.

We also give errors from the quarterly models given in the previous Section, §8.4.

	Q 1	Q 2	Q3
Actual NFC	464	444	495
Error of predictions			
Sum of monthly TSM prediction (Table 2.1d)	7	- 6	8
Reduced structural model, $X = FUEL$ , FUEL predicted by previous year's value in same quarter (Table 8.4d)	- 21	- 3	- 19
Reduced structural model, $X = FUEL$ , FUEL predicted by TSM (Table 8.7b)	- 32	- 43	- 23

We note that the quarterly errors for the monthly pure time series model are smaller than those for the two predictions involving the quarterly reduced structural model and FUEL. Theoretically one would expect better forecasts to be derived from monthly data than quarterly data because the quarterly forecasts could be derived as special cases of the monthly data.

Here we investigate further structural models for monthly data and fuel. We recall from §6.1 that the BATS analysis for AUSACCM (Table 6.1) suggested that the model with FUEL and LAG FUEL was not a good one although the reduced structural model analysis of Table 6.2b had both FUEL and LAG FUEL statistically significant.

The BATS analysis, we remind the reader from §5, allows for both a stochastically evolving trend and seasonal pattern for the dependent variable. If an explanatory variable is following largely the same stochastic evolution as the dependent variable then it will have little explanatory power and the regression coefficient will not be significantly different from zero.

In Figures 8.1 and 8.2 we give on-line estimates (estimate given by the the solid line, 90% confidence limits by the dashed lines) of the regression coefficients for FUEL and LAG FUEL from the BATS analysis. If the reduced structural model with estimates  $(2.16 \times 10^{-6} \text{ and } 3.16 \times 10^{-6})$  for FUEL and LAG FUEL were adequate, then the plots in Figures 8.1 and 8.2 should be horizontal at the values 2.16 x  $10^{-6}$  and 3.16 x  $10^{-6}$  respectively. This is obviously not the case. We summarise the situation below.

	FUEL estimate x 10 <sup>6</sup>	LAG FUEL estimate x 10 <sup>6</sup>
Reduced structural model 2	2.16	3.16
	(1.02)	(1.02)
	Z = 2.12	Z = 3.10
BATS analysis	- 2 in 1982	- 1 in 1982
•	0 in 1987	varying about 0
	1 in 6/89	1984 to 1991
	to 6/90	

The BATS analysis therefore suggests little explanatory power from fuel sales as the on-line estimated regression coefficients are never significantly different from zero.

The model with FUEL for monthly data needs further investigation. Preliminary results when fitting the reduced structural model 2 with FUEL, LAG FUEL and LAG LAG

FUEL to the revised data up to Q3 1991 give statistically significant coefficients for FUEL and LAG FUEL, equal to about  $3 \times 10^{-6}$ , and nonsignificant coefficient for LAG LAG FUEL.

This model suggests therefore that some account should be made for the delay in retail sales of vehicle fuel as compared with wholesale sales of fuel, as measured by FUEL, which could be up to two months, but unlikely to be longer.

#### 8.8 Conclusion

The Reduced Structural model 2 for quarterly Australian NFC data involving FUEL

appears to give a reasonable fit to the data ( $R_S^2 = 0.48$ ) with a statistically significant

(Z = 4.35) regression coefficient estimated to be 2.78 x 10<sup>-4</sup> accidents per quarter per kilolitres of fuel; Table 8.4a. The model can be used to 'explain' the extent to which the decrease in crash numbers in 1990 and 1991 are due to economic factors as accounted for in the model by FUEL. Table 8.4c gives predictions of quarterly NFC for 1990 using the data up to Q4 1989 but the actual FUEL values for 1990. These predictions are those which the model gives by taking into account actual economic conditions. We can compare these with predictions made using 1989 FUEL values used for 1990 FUEL, that is a projection of 1989 economic conditions, as given by FUEL into 1990. Below we give the predictions made using reduced Structural model 2.

#### **Actual and Predictions for 1990**

	Q1	Q2	Q3	Q4	Total
Actual	498	510	544	496	2048
*Actual FUEL	561	548	588	506	2203
+ 1989 FUEL	557	545	561	580	2242

#### 1989 actual

2402

\*Table 8.4c; + new predictions.

We can express the totals for each row as a percentage of 1989 actual NFC:

1990 actual	85.3
Prediction using 1990 FUEL	91.7
Prediction using 1989 FUEL	93.4
1989 actual	100%

Thus the decrease in NFC seen in 1990 compared with 1989, 100 - 93.4 or 6.6%, could be attributed to continuing trends of road safety; a further 93.4 - 91.7, or 1.7%, attributed to the economic conditions of 1990 and 91.7 - 85.3 or 6.4% is unexplained by the model. The 6.4% difference could be due to road safety measures unaccounted for, unexplained economic effects or other factors.

Carrying out a similar analysis for the first three quarters of 1991 we obtain these figures.

#### Actual and Predictions for 1991

	Q1	Q2	Q3	Total
Actual *Actual FUEL	464 446	444 445	495 453	1403 1344
+ 1989 FUEL	485	477	514	1476

#### 1989 Total

1786

\*Table 8.4e; + Table 8.4d.

We can now express these totals as a percentage

1991 actual	78.6%
Prediction using 1991 FUEL	75.3%
Prediction using 1989 FUEL	82.6%
1989 actual	100%.

In this particular case the model predictions using actual FUEL under predict. The predictions using 1989 FUEL values for both 1990 and 1991 over predict the 1991 actual value by 4%. In this the interpretation is not as straightforward, but the model suggests that the 1991 total is as bad as to be expected but better than if 1989 economic conditions had continued. It could be argued from the model that the recession of 1990 and 1991 has saved lives by reducing NFC by (2402 - 2203) or 199 crashes in 1990 and (1786 - 1344) or 442 crashes in the first 3 quarters of 1991 - a total of 641 crashes. This and other factors have reduced the number of fatal crashes by, in total, 737 crashes.

Finally we can consider predictions for 1991 using data up to the end of 1989, FUEL for 1990 and 1991 equal to the values of FUEL for 1989. These predictions for 1991 represent continuing economic conditions of 1989 and road safety trends of 1989. The predictions are

Q1	Q2	Q3	Total
516	505	520	1541

as a percentage of 1989 figure, the total, 1541, is 86.3% whereas, as above, the actual for 1991 is 1403 or 78.6% of the 1989 corresponding total.

We now have various predictions for 1991 (Q1 - Q3) as follows

Base 1989, FUEL $=$	1989	86.3%
Base 1990, FUEL =	1989	82.6%
Base 1990, FUEL =	actual	75.3%

# Actual 78.6%.

The drop 100 - 86.3 or 13.7% is what might have been expected given continuing economic activity and road safety trends from 1989. The further drop 86.3 - 82.6 or 3.7% could represent new road safety trends based on 1990 trends. The difference 82.6 - 75.3 or 7.3% might be the further drop which would have been expected due to actual economic conditions. In reality, this is an overestimate of the improvement seen in 1991. In all of these predictions, however, the overwhelming message is the size of the effects due to economic factors as represented by FUEL.

In summary, we have found models which are both useful for explanation and prediction. These include for explanation the following.

- . Australian quarterly data a model involving fuel sales.
- Australian monthly data a model involving current and lagged fuel sales; this model requires further investigation.
- . Victoria quarterly data a model involving fuel sales.
- **NSW and Queensland monthly data** a monthly involving a weather index.

For prediction time series models for Australian and States monthly data perform well.

#### Table8.1

#### Summary of Monthly & Quarterly Best Models (fitted to data up to end 1990)

#### Monthly Data

	Explanatory Variables	Structural Model	Regression Estimate	Z-value of estimated parameter	R <sup>2</sup> S
Australia	FUEL + LAG FUI	2 EL	2.2 x 10 <sup>-6</sup> 3.2 x 10 <sup>-6</sup>	2.1 3.1	0.510
NSW	MVR	2	6.8 x 10 <sup>-6</sup>	0.5	0.396
	WI	2	- 1.0 x 10 <sup>-2</sup>	- 1.2	0.445
Victoria					
	MVR WI	2, 5 3	3.9 x 10 <sup>-6</sup> -2.7 x 10-3	0.3 - 0.4	0.363 0.364
Queensland	!				
	WI	2	- 1.1 x 10 <sup>-2</sup>	- 1.9	0.387

.

•

	Explanatory	Structural	Regression	Z-value	$R_{s}^{2}$
	Variables	Model	Estimate	of estimated parameter	5
Australia					
	FUEL	2	1.8 x 10 <sup>-8</sup>	3.5	0.526
	FUEL +	2	1.8 x 10 <sup>-8</sup>	3.1	0.363
	%CHGP	ET	- 9.2 x 10 <sup>-3</sup>	- 0.4	
Victoria					
	FUEL	2	7.8 x 10 <sup>-8</sup>	2.4	0.639
	FUEL +	5	7.8 x 10 <sup>-8</sup>	2.2	0.639
	%CHGP	ET	- 3.0 x 10 <sup>-4</sup>	- 0.35	

# Table 8.2

Detailed Results for the Australian Quarterly Series AUSACCQ with FUEL as explanatory variable - Reduce Structural Models 1 and 2 (fitted using data up to 4th Quarter, 1989)

#### Table 8.2a

## Model Estimates

	SM1		SM2		
Term	ESTIMATE (standard error)		ESTIMATE (standard error)		
trend Q2 Q3 Q4 FUEL MA(1) IMA(1) R <sup>2</sup>	-8.4 x 10 <sup>-4</sup> -0.00171 -0.00020 -0.00093 2.8 x 10 <sup>-8</sup> -0.708 0.86	(8.3 x 10 <sup>-5</sup> ) (0.000886) (0.00114) (0.00110) (3.7 x 10 <sup>-9</sup> ) (0.148)	-0.000818 -0.001212 0.000315 0.00034 2.2 x 10 <sup>-8</sup> 0.425	(0.000234) (0.000909) (0.000906) (0.00130) (5.6 x 10 <sup>-9</sup> ) (0.175)	
$R_{s}^{2}$	0.65		0.54		

# Table 8.2b

<b>Predictions</b> for	1990 based on v	values of AUSAC	CQ up to Q4,	1989 and FUEI	up to
		Q4, 1990.			_

1990 Quarter SM 1	Forecast	95% Confidence Limits Limits	Actual	Error
Q1	0.0331	0.0300, 0.0361	0.029	-0.004
Q2	0.0332	0.0294, 0.0370	0.030	-0.003
Q3	0.0364	0.0326, 0.0401	0.032	-0.004
Q4	0.0284	0.0246, 0.0321		
SM 2				
Q1	0.0324	0.0289, 0.0359	0.029	-0.003
Q2	0.0313	0.0273, 0.0353	0.030	-0.001
Ō3	0.0340	0.0295, 0.0385	0.032	-0.002
Q4	0.0282	0.0232, 0.0331		

#### Table 8.2c

# Summary Values up to Q4 1989

Average value of AUSACCQ	0.0409
Minimum value of AUSACCQ	0.0340
Maximum value of AUSACCQ	0.0510
Average value of FUEL	3.97 x 10 <sup>6</sup>
Minimum value of FUEL	3.61 x 10 <sup>6</sup>
Maximum value of FUEL	4.39 x 10 <sup>6</sup>

#### Table 8.3

# Detailed Results for the Australian Quarterly Series AUSACCQ with FUEL as explanatory variable - reduced structural models 1 and 2 (fitted using data up to 4th Quarter, 1990)

#### Table 8.3a

#### SM1

### SM2

Term	ESTIMA (standard	ESTIMATE (standard error)			ATE error)
trend Q2 Q3 Q4 FUEL MA(1) R <sup>2</sup>	-0.000857 -0.001389 0.00005 -0.00031 2.6 x 10 <sup>-8</sup> -0.777 0.88	(0.000080) (0.000873) (0.00117) (0.00105) (3.3 x 10 <sup>-9</sup> ) (0.124)	IMA(1)	-0.000854 -0.000868 0.000641 0.00083 2.1 x 10 <sup>-8</sup> 0.407 0.87	(0.000228) (0.000833) (0.000854) (0.00124) (5.4 x 10 <sup>-9</sup> ) (0.166)
$R_s^2$	0.53			0.46	

#### Table 8.3b

#### **Predictions** for 1991. FUEL for 1991 is taken as same as for 1990.

1991 Quarter SM1	Forecast	95% Confidence Limits Limits	Actual	Error
Q1 Q2 Q3 Q4	0.02967 0.02853 0.03142 0.02431	0.02644, 0.03290 0.02444, 0.03262 0.02733, 0.03551 0.02022, 0.02840	0.027 0.026	-0.003 -0.003
SM 2				
Q1 Q2 Q3 Q4	0.02697 0.02616 0.02871 0.02321	0.02352, 0.03043 0.02214, 0.03017 0.02420, 0.03321 0.01826, 0.02816	0.027 0.026	0.00 0.00

# Table 8.3c

# Detailed Results for the AUSACCQ series with FUEL as explanatory variable - reduced structural models 1 and 2 (fitted using data up to 3rd Quarter, 1991)

#### SM1

#### SM2

Term	ESTIMATE (standard error)	ESTIMATE (standard error)	
trend Q2 Q3	-0.000780 (0.000064) -0.001098 (0.000786) 0.00036 (0.00103)	-0.000762 (0.000207) -0.000764 (0.000735) 0.000706 (0.000781)	
Q4 FUEL R <sup>2</sup>	0.000636  (0.000927) 2.2 x 10 <sup>-8</sup> (3.2 x 10 <sup>-9</sup> ) 0.91	$1.8 \times 10^{-8} (4.1 \times 10^{-9})$ 0.90	
$R_s^2$	0.56	0.50	

# Table 8.4

- - - - --- ----

# Reduced Structural model estimates for raw NFC quarterly data (fitted using data up to 3rd Quarter 1991)

Table 8.4a						
Structural Model						
	1	2	3	4	5	
FUEL (estimate x 10 <sup>5</sup> )	34.3	27.8	27.8	27.7	27.8	
(std error x 10 <sup>5</sup> )	(4.64)	(6.39)	(6.48)	(6.84)	(6.50)	
$R_{S}^{2}$	0.54	0.48	0.47	0.42	0.47	

## Table 8.4b

## Detailed Results for Models 1 and 2

	SN	11	SN	/12
Term	ESTIMATE (standard error)		ESTIMATE (standard err	
trend MA(1)	- 10.04 -0.666	0.95 (0.130)	- 9.47	3.15
IMA(1)		(01120)	0.393	(0.160)
R <sup>2</sup>	0.87		0.86	
R <sub>s</sub> <sup>2</sup>	0.54		0.48	

#### Table 8.4c

# Predictions for 1990 based on values of raw NFC quarterly up to Q4, 1989 and FUEL up to Q4, 1990.

1990 Quarter	Forecast	95% Confidence Limits	Actual	Error
SM1				
Q1	570.3	523.1, 617.4	498	-72
Q2	573.2	515.4, 631.1	510	-63
Q3	621.4	563.6, 679.3	544	-77
Q4	506.5	448.6, 564.3	496	-11
SM1				
Q1	561.4	507.8, 615.0	498	-63
Q2	547.7	486.0, 609.4	510	-38
<b>Q</b> 3	588.4	519.5, 657.2	544	-44
Q4	506.4	431.0, 581.7	496	-10

#### Table 8.4d

# Predictions for 1991 based on values of raw NFC quarterly data up to Q4, 1990. FUEL for 1991 is taken as the same as for 1989, quarter by quarter.

1991 Quarter	Forecast	95% Confidence Limits	Actual	Error
SMI				
<b>O</b> 1	536.8	486.0, 587.6	464	-73
òż	513.4	452.3, 574.4	444	-69
Õ3	554.5	493.5, 615.6	495	-60
Q4	470.8	409.8, 531.9		
SM2				
<b>O</b> 1	485.4	432.0, 538.9	464	-21
õ2	476.9	414.2, 539.6	444	-33
Ō3	513.6	442.9, 584.4	495	-19
Q4	449.5	371.6, 527.5		

## Table 8.4e

# Predictions for Q1, Q2, Q3 1991 based on values of raw NFC quarterly data up to Q4 1990 and actual FUEL up to Q3, 1991

1991 Quarter	Forecast	95% Confidence Limits	Actual	Error
SM1				
01	488.9	438.0, 539.7	464	-25
òz	473.9	412.9, 535.0	444	-30
Q3	480.1	419.0, 541.1	495	15
SM2				
01	446.1	392.7, 499.6	464	18
Õ2	444.6	381.9, 507.3	444	- 1
Q3	452.5	381.7, 523.2	495	42

#### Table 8.5

# ARIMA Time Series Models for FUEL, quarterly series from Q2 1981 until Q4 1990

FUEL Model 1

ł

#### SARIMA (1111004)

<b>AR</b> (1)	-0.381	(0.248)
SAR(4)	0.569	(0.150)
MA(1)	0.455	(0.224)

Mean Square Error =  $8.27 \times 10^9$  on 35 degrees of freedom.

Chi-square statistic for residual autocorrelations = 7.437 on 19 degrees of freedom.

FUEL Model 2	SARIMA (0110014)
MA(1)	0.646 (0.132)
SMA(4)	-0.486 (0.170)

Mean Square Error 9.73 x  $10^9$  on 36 degrees of freedom.

Chi-square statistic for residual autocorrelations = 15.747 on 19 degrees of freedom.

FUEL Model 3	SARIMA (0111004)
<b>MA</b> (1)	0.698 (0.135)
SAR(4)	0.626 (0.146)

Mean Square Error =  $8.46 \times 10^9$  in 36 degrees of freedom.

Chi-square statistic for residual autocorrelation = 8.775 on 17 degrees of freedom.

#### Table 8.6

#### Predictions for FUEL for 1991 using models of Table 8.5 Model fitted to data up to end of 1990 (Units millions of kilo litres)

#### FUEL Model 2 (of Table 8.5) SARIMA (0110014)

1991 Quarter	Forecast	95% Confidence Limits	Actual	Error
Q1	4.249	4.049, 4.450	4.122	-0.127
Õ2	4.253	4.040, 4.465	4.189	-0.064
Õ3	4.299	4.076, 4.523	4.177	-0.122
Q4	4.175	3.941, 4.410		

#### FUEL Model 3 (of Table 8.5) SARIMA (0111004)

1991 Quarter	Forecast	95% Confidence Limits	Actual	Error
Q1	4.253	4.066, 4.439	4.122	-0.131
Õ2	4.280	4.085, 4.474	4.189	-0.091
<b>O</b> 3	4.335	4.132, 4.538	4.177	-0.158
Q4	4.193	3.982, 4.404		

#### Table 8.7

# Predictions for AUSACCQ and raw quarterly NFC using reduced structural model 2 with X = FUEL and FUEL predicted by a TSM

#### Table 8.7a

#### Predictions for AUSACCQ for 1991 using data up to the end of 1990

1991	Forecasts	95% Confidence Limits*
Q1	0.0284	(0.0245, 0.0314)
Q2	0.0273	(0.0225, 0.0316)
Q3	0.0291	(0.0237, 0.0344)
Q4	0.0260	(0.0204, 0.0327)

#### Table 8.7b

#### Predictions for raw quarterly NFC for 1991 using data upto the end of 1990

1991 Quarter	Forecast	95% Confidence Limits*	Actual	Error
01	496.4	(440.6, 552.2)	464	-32
ò	486.6	(415.7, 557.5)	444	-43
Q3	516.6	(433.3, 599.9)	495	-22
04	488.4	(394.3, 582.5)		

• Confidence intervals include errors of prediction for FUEL.



ON-LINE COEFFICIENT OF FUEL

Figure 8.1



ON-LINE COEFFICIENT OF LAGFUEL

Figure 8.2
#### PART 3

#### **Fatalities Per Crash**

#### Overview

A simple time series model is suggested for the variable fatalities per crash. Also, the ACT is found to have an estimate of fatalities per crash which is statistically different from other jurisdictions.

#### 9. MODELLING OF THE NUMBER OF FATALITIES PER CRASH

Data for the number of crash fatalities (NCF) were investigated for the period April 1975 until December 1990 on a monthly basis for Australia. Initially, the variable Fatalities per Crash (FC) was derived by dividing NCF by NFC for each month; summary statistics are given in Table 9.1, indicating an average of 1.129 fatalities per crash, with an estimated standard error of 0.003. The summary indicates the variation is quite large, with the lower and upper quartiles given by 1.104 and 1.148 and minimum and maximum values given by 1.057 and 1.286. We have not standardised the number of accidents or fatalities by days in the month or population because the effect would be very small for these analyses, being a second order effect of variance rather than first order of mean.

An alternative estimate of the average fatalities per crash is given by a regression ratio estimate. This turns out to have the same precision (standard error 0.003) as the monthly average of FC values found above. This analysis also investigates differences between Australia, states and territories. Estimates found by regressing, with no intercept, the number of crash fatalities (NCF) per month on the number of fatal crashes (NFC) per month for Australia, the states and territories were obtained and estimates are given in Table 2. Without making any allowance for multiple comparisons, the Z value for comparing the largest value of the states and territories (except ACT) with the smallest (NSW, 1.136 against Tasmania,1.113) is 2.1, just statistically significant at the 5 percent level. It would probably be reasonable to assume no or little differences between states and territories. The ACT does have a rate which is statistically significantly different from the other states and territories.

From the plot of Australia FC series against time, Appendix D Figure D1, it appears that there might be a seasonal or monthly effect. The Seasonal Subseries plot, Appendix D, Figure D2, confirms a strong seasonal (monthly) effect. The Seasonal Subseries plot depicts the 12 monthly averages as horizontal lines with values from succeeding years for the given month plotted as vertical lines to the horizontal line. From the plot, there are local peaks in the monthly effects at May, September and December. There is no apparent trend in the yearly values since the vertical lines are haphazardly patterned. A Seasonal Subseries plot is given for the series of seasonal (monthly lag 12) differences, Appendix D, Figure D3, which shows little pattern whatsoever and suggests purely monthly/seasonal effects. An ANOVA confirms monthly effects for the Australian data. Results are given in Table 3 where an ANOVA for between and within months is given both for the original series and the seasonally differenced series. There are significant differences between month effects which are indicated by the significance of the between month effect for the original data and the insignificance of the between month effect for seasonally differenced data.

In Table 9.4 we give the monthly means for the series, giving the high months as December (significantly different from an average month) and January (not significantly different) and the low month as July (significantly different from an average month). A

simple prediction method would be to use the appropriate monthly average for July and December and the average, 1.129 (which remains unchanged), for the remaining months, or extend this method to include February and June, which have marginally significant effects, and the average of the remaining months, 1.134. This suggests that a time series approach to predicting the number of fatalities per fatal accident may be worthwhile. A reduced structural model 2 (see section 5) with explicit trend and monthly components was fitted to the FC series. The moving average error of order 1 was, however, replaced by moving average error of order 12, and the lag 6 and 11 terms found to be significatly different from zero. All other moving average terms were not significatly different from zero. Although a time series approach is possible to implement to provide forecasts, it is doubtful whether such forecasts would be an improvement over monthly means.

#### Table 9.1

#### Summary statistics for Fatalities per crash: Australia April 1975 to December 1989

Average	1.129 (0.003)
Median	1.127
Standard Devn	0.037
Min, max	1.057, 1.286
Lower, upper quartiles	1.104, 1.148

#### Table 9.2

#### Fatalities per Fatal Crash, monthly data. Results from Regression Analysis

#### jurisdiction

estimate (s.e.)

Australia	1.130 (0.003)
NSW	1.136 (0.004)
Victoria	1.127 (0.005)
Queensland	1.133 (0.005)
South Australia	1.120 (0.006)
WA	1.124 (0.008)
Tasmania	1.113 (0.010)
Northern Territory	1.131 (0.015)
ACT	1.081 (0.013)

#### Table 9.3

#### ANOVA for monthly series for number of fatalities per crash

#### **Original Data**

Source	Sum o	of Squares	Df	F statistic
Between years Between months	0.0304		14 11	2.05
Within months and years	0.2332		154	5.56
Seasonally Differenced	Data			
Between years	0.0550		14	1.79
Between months Within months and years	0.0010 0.3316	151(3)	11	0.04

#### Table 9.4

### Monthly means for the number of fatalities per crash and, in parentheses, the difference from the average 1.129 multiplied by 100

January	1.147 (18)	July	1.104	(-25)
February	1.119 (-20)	August	1.124	(-5)
March	1.116 (-13)	September	1.139	(10)
April	1.127 (-2)	October	1.133	(4)
May	1.136 (7)	November	1.126	(-3)
June	1.109 (-20)	December	1.158	(29)

(standard error of difference of monthly means, 0.012).

#### References

Box, G.E.P. and Jenkins, G.M. (1970) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Box, G.E.P. and Pierce, D.A. (1970) Distribution of the Residual Autocorrelations. J.Am.Statist.Assoc., 65, 1509 - 26.

Donaldson, P.K., Gillan, P. and Jones, B.P. Road Transport demand in Australia. Conf. of Economists, September 1990, Sydney.

Harvey, A.C. (1989) Forecasting, structural time series models and the Kalman filler. Cambridge: Cambridge University Press.

Harvey, A.C. and Durbin, J. (1986) The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. J. Roy. Statist.Soc. A 149, 187 - 227.

Ljung, G.M. and Box G.E.P. (1978) On a measure of lack of fit in time series models. Biometrika, **66**, 67 - 72.

McCullagh, P. and Nelder, J. (1989) Generalized linear models, 2nd edition, London: Chapman & Hall.

Payne, R.W. et al (1987) GENSTAT 5 Reference Manual. Oxford: Clarendan Press.

STATGRAPHICS (1989) Statistical Graphics System Version 4.0. STSC Inc.

West, M. and Harrison, J. (1989) Bayesian Forecasting and Dynamic Models. New York: Springer Verlag.

West, M., Harrison, J. and Pole, A. (1987) BATS: A user guide. Warwick Research Report 114, Dept. of Statistics, University of Warwick.

### VOLUME 2 APPENDICES

# *IMPORTANT*

## PLEASE DO NOT STAMP OR STAPLE THIS DOCUMENT

### APPENDIX A

### Tables and Figures for Monthly and Quarterly series of number of fatal road crashes and explanatory variables.

AUSTRALIA MONTHLY	DATA	
TABLE A.1. Number	of fatal road crashes by da	y

ROW	YEAR	JAN	FEE	3 ⊨	IAR	APR		MAY	JUN	JUL
1	1976	8 0323	7 8964	< 835	48	9 1000	8	7419	8 2667	7 5806
2	1077	7 4516	R 1/20	0.00	59	9 9000	9. 9	5904	0.2007	9 1075
7	1079	0 75/0	0.1423	, , , ,		7 0447	· · ·	J000 E/d/	0.3333	0.1733
2	1070	4 0072	7 71/3	I 9.0/ I 0.4/	/4 50	7.9007		3404 90/F	9.0007	9.1290
4	17/7	0.9032	7.7143	, 9.04 7 7 07	52	9.33333		6000	0.200/	0.4010
2	1980	0.7742	0.3103	) (.Y3	22	8.8000	<u> </u>	6129	8.5000	8.6//4
<u>0</u>	1981	7.7419	(.5/14	• (.2)	205	8.2000	· · ·	8387	7.8000	8.1935
	1982	8.0645	8.3571	6,90	32	8.4333	8.	0323	7.5667	7.8065
8	1983	5,8065	5.9286	5 6.87	10	6.7333	6.	6129	7.1667	6.8387
9	1984	6.5484	6.2069	7,25	81	6.5333	6.	5806	6.5667	5.4194
10	1985	6.1935	7.3214	6.83	87	6.7667	7.	6452	7.0333	6.9032
11	1986	6.7742	7.2500	) 7.93	55	7.6333	7.3	3226	6.8667	6.2903
12	1987	6.6452	5.6429	7.06	45	6.4667	6.	3548	6.5667	7.2581
13	1988	6.7097	7.2143	5 7.54	84	6.1000	7.	1613	6.1667	6.6129
14	1989	6.1290	6.7857	7 7.25	81	5.5333	6.	1290	7.2000	6.1613
15	1990	4.9355	5,2143	5 6.41	94	5.3000	5.	2581	6.2667	5.9677
ROW	YEAR	AUG	SE	P	ОСТ	NOV	1	DEC		
1	1976	8.2258	9.633	33 8.7	419	8.6667	7 1	0.1290		
2	1977	8.7742	8.100	0 8.5	032	9.5333	5	9.8065		
3	1978	10.0645	8.900	0 9.1	613	9.0333	5	8.8387		
4	1979	8.8065	8.866	57 8.3	871	7.7000	) 1	0.1613		
5	1980	8,8065	7.633	33 8.3	548	9.5000	)	7.9677		
6	1981	7.3226	8.766	57 8.5	806	8.4333	5	7.9355		
7	1982	8.2581	7.500	0 8.1	290	7.9333	5	7.5161		
8	1983	6.9355	7.733	33 6.4	516	7.1333	5	7.0645		
9	1984	7.0968	7 437	3 6.7	419	7.4000		8 2581		
10	1985	5 9677	7 464	57 7 4	516	7 7000		8 9677		
11	1086	6 6452	6 500	10 7 C	000	7 1333	i i	7 2581		
12	1087	6 06/5	7 000	10 7 4	170	7 6000	Ś	7 35/8		
17	1000	7 2003	8 044	10 7.C	971	4 0447	,	7 0000		
17	1700	5 8710	7 944	27 7 2 7	0071	4 100/		7 4776		
14	1000	5.0/10	(,000		903	5.1000	,	1.0/14		
TABLE	A.2. Ne	w Motor Vehic	le Regis	stration	/42 IS	5.2000	,	0.1933		
DOLL	VEAD					<b>DD</b>				
KOW	TEAK	JAN	FEB	MAK	A	PK	MAT	101	I JUL	
1	1981	*	*	51059	489	48 5°	1508	56222	2 59278	
2	1982	43847	45002	59485	501	11 54	6043	59146	55565	
3	1983	34924	40147	56966	422	16 49	2903	50945	5 45697	,
4	1984	41610	49612	57167	462	77 60	0993	57552	2 57603	
5	1985	50568	56284	64435	565	76 64	1554	57661	61673	
6	1986	43481	41259	44083	519	55 47	7354	44798	3 48787	,
7	1987	33330	34464	39327	320	15 34	5773	40674	40335	
Ŕ	1088	20686	35473	42903	367	45 44	335	46448	3 40240	
0	1080	35/35	45052	52091	/30	87 53	717	56104	51427	
10	1990	44212	48213	67172	472	84 60	353	52463	55863	
ROW	YEAR	AUG	SEP	ост	N	ov	DEC			
1	1981	51727	50406	46963	488	00 53	3434			
2	1982	56500	50587	47940	467	63 57	7567			
3	1983	54268	46768	44284	499	46 49	9540			
4	1984	56884	49256	56770	536	31 50	8060			
5	1985	56194	54547	57599	587	03 50	0568			
6	1986	48119	42137	40510	355	65 38	3091			
7	1987	35308	40672	38947	354	17 44	551			
Ŕ	1988	46187	50071	46066	529	80 44	2405			
ŏ	1980	54790	55232	51183	532	65 47	7168			
10	1000	50691	50261	53345	474	84 43	204			
10	1770	20001	20201							

•

NSW MONTHLY DATA TABLE A.3 Number of fatal road crashes per day

ļ

ROW	YEAR	JAN	FEB	MAR	APR	MAY	JUN
1	1074	2 77/10	2 / 1370	2 7/10/	Z ZZZZZ	7 51617	3 00000
<b>.</b>	1077	2.//417	2.41377	7 10755	7 74447	3.31013	3.00000
4	1977	2.04310	2.90429	3.19333	3.30007	2.93340	2.3000/
Ś	1978	3.10129	3.21429	3.0/742	2.93333	3.29032	3.90000
4	1979	2.09677	2.50000	3.61290	3.73533	3.29032	2.73333
5	1980	2.77419	1.93103	2.80645	3.66667	3.06452	3.26667
6	1981	3.12903	3.03571	2.54839	3.60000	2.93548	2.53333
7	1982	3.74194	3.50000	2.67742	3.36667	2.93548	2.93333
8	1983	2,09677	1.71429	2.58065	2.63333	2 19355	2.50000
ō	1984	2 54830	2 34483	2 77419	2 13333	2 45161	2 56667
10	1085	2 00677	2 80784	2 04452	2 54447	2 4/514	2 20000
11	1084	2.07077	2.07200	2.00472	2.00007	2.04310	2.30000
10	1900	2.4030/	2.00/14	2.0/09/	2.40000	5.00000	2.23333
12	1987	2.51615	1.75000	2.38/10	2.50000	1.96774	2.50000
13	1988	2.51613	2.39286	2.77419	2.20000	2.35484	2,30000
14	1989	1,54839	2.64286	2.32258	1.33333	2.09677	2.50000
15	1990	1.38710	1.78571	2.48387	1.73333	1.74194	1.96667
POU	VEAD		AUC	SED.	007	101	050
ROW	TEAK	305	AUG	357	ΨLI	NUV	UEC
1	1976	2.67742	2.90323	3.83333	3.35484	2.86667	3.25806
2	1977	3.00000	3,19355	2.73333	3.48387	3.30000	3.35484
3	1978	3.64516	3.61290	3.33333	2,93548	3.06667	3.38710
4	1979	3,16129	2 90323	3,40000	3.03226	2 96667	3 51613
5	1980	3 7/ 19/	3 58065	2 80000	3 35/.8/	3 93337	2 00323
,	1081	3 14120	2 77/10	3 10000	7 / 9797	7 / 4 4 4 7	7 79710
7	1001	7 00/77	2.11417	3.10000	3.4030/	3.40007	3.30710
1	1982	3.090//	3.090//	2.90007	5.06452	3.00000	2.32258
8	1983	2.32258	2.55484	2.66667	2.35484	2.60000	2.77419
9	1984	1.77419	2.58065	2.50000	2.22581	2.83333	3.12903
10	1985	2.41935	2.35484	2.86667	2.77419	2.90000	3.51613
11	1986	2.22581	2.06452	2.30000	2.83871	2.33333	2.54839
12	1987	2.41935	1.83871	2,50000	3.09677	2.33333	2.32258
13	1988	2 45161	2 16129	2 73333	3 09677	2 36667	2 70968
16	1080	2 06452	1 80445	2 80000	2 16120	2 10000	2 / 1075
15	1000	2.00452	1.00727	2.00000	1 70049	2.10000	2.41733
	1770	2.41733	1.90323	2.30000	1.70900	1.00007	2.03220
TABLE	A.4 Weath	er Index					
BOU	VEAD						
KOM	TEAK	JAN	FEB	MAK	APR	MAY	JUN
1	1981	*	*	5.1170	6.9606	12.0703	7.9067
2	1982	9.2734	7.2031	16.3674	8.9527	3,9842	15.7568
ž	1983	5 3910	10 8824	9 1170	12 3595	12 0207	10 2188
ž	108/	1/ 07/3	17 8350	10 2655	0 1052	10 0151	5 4454
5	1095	7 0170	10 2071	11 2/77	15 700/	17 0757	7 70/0
2	1902	2.9139	10.2031	11.04//	12.7004	17.9/37	7.3049
D	1900	10.9606	10.8508	1.18/3	7.8042	11.0467	3.1045
(	1987	8.4692	8.8120	11.4555	7.1873	10.1406	10.0782
8	1988	9.0624	11.1012	13.0618	17.1952	14.0703	11.8981
9	1989	18.8035	10.8429	18.5459	20.7489	17.8824	18.0388
10	1990	11.2031	17.6786	17.8508	15.2583	14.0309	12.0624
POU	YEAR	.un	AUG	SEP	ост	NOV	DEC
NO.		002			001		020
1	1981	4.9067	6.1098	4.1564	11.9060	14.2110	9.3595
2	1982	13.3983	4.1249	10.1249	11.9369	3.9842	8.0940
3	1983	10.2425	10.8902	9.0000	12.9527	11.3207	13.8745
4	1984	16.0861	7.7812	10.0940	11.2188	12.8666	10,1091
5	1985	8,1170	10, 1643	11.2734	15.7884	10.4692	12.9763
Ä	1984	Q 0315	12.0703	11.0388	9,0237	13 7417	6 0040
7	1097	11 0404	16 12/0	1 0005	16 273/	6 0/55	10 09/2
	1707	11.7000	10.1247	4.7761	14.2/34	0.7400	10.7042
~ ~	1020	2 2027	<u> </u>	10 0120			17 2021
ž	1988	6.2037	9.2425	10.9139	0.1019	12.0010	17.0000
9	1988 1989	6.2037 8.4384	9.2425	10.9139	5.1564	12.5616	12.9606

ı.

VICTORIA MONTHLY DATA TABLE A.5 Number of fatal road crashes by day

RO₩	YEAR	JAN	FEB	MAR	APR	MAY	JUN
1	1976	2.16129	2.31034	2.45161	2.40000	2.06452	2.43333
2	1977	1.80645	2.32143	2.61290	2.20000	1,90323	2.43333
3	1978	2.00000	1.89286	2.77419	1.90000	1,93548	2.13333
4	1 <b>979</b>	1.93548	1.71429	2.54839	2.16667	1.41935	2.00000
5	1980	1.41935	1.79310	1.96774	1.90000	1.45161	1.76667
6	1981	1.77419	1.96429	1.70968	1.70000	2.03226	2.03333
7	1982	1.61290	1.64286	1.51613	1,80000	1.96774	1.43333
8	1983	1.03226	1.67857	1.58065	2,00000	1.74194	1.76667
9	1984	1.90323	1.58621	1.77419	1.43333	1.58065	1.43333
10	1985	1.45161	1.50000	1.48387	1.86667	2.03226	1.53333
11	1986	1.58065	1.64286	2.00000	1.90000	1.77419	1.50000
12	1987	1.67742	1.32143	1.77419	1.56667	1.80645	1.83333
13	1988	1.41935	1.92857	2.06452	1.66667	2.22581	1.56667
14	1989	1.96774	2.14286	2.35484	1.80000	1.96774	2.10000
15	1990	1.41935	1.32143	1.67742	1.20000	1.64516	1.70000
ROW	YEAR	JUL	AUG	SEP	ост	NOV	DEC
1	1976	1.83871	2.03226	2,23333	2.16129	2.30000	2.74194
2	1977	2.09677	2.51613	2.10000	1.87097	2,63333	2.93548
3	1978	2,12903	2.35484	1.96667	2.32258	2,20000	1.90323
4	1979	2.12903	2.19355	1.86667	2.19355	2.00000	2.54839
5	1980	1.77419	1.54839	1.43333	1.64516	1.80000	1,51613
6	1981	1.83871	1.77419	2.06667	1.96774	1.73333	1.54839
7							
•	1982	1.41935	2.00000	1.66667	1.90323	1.90000	1,90323
8	1982 1983	1.41935 2.00000	2.00000 1.54839	1.66667 1.83333	1.90323 1.32258	1.90000 1.63333	1.90323 1.54839
8 9	1982 1983 1984	1.41935 2.00000 1.29032	2.00000 1.54839 1.41935	1.66667 1.83333 1.76667	1.90323 1.32258 1.41935	1.90000 1.63333 1.56667	1.90323 1.54839 1.74194
8 9 10	1982 1983 1984 1985	1.41935 2.00000 1.29032 1.58065	2.00000 1.54839 1.41935 1.45161	1.66667 1.83333 1.76667 1.56667	1.90323 1.32258 1.41935 1.67742	1.90000 1.63333 1.56667 1.70000	1.90323 1.54839 1.74194 1.87097
8 9 10 11	1982 1983 1984 1985 1986	1.41935 2.00000 1.29032 1.58065 1.29032	2.00000 1.54839 1.41935 1.45161 1.64516	1.66667 1.83333 1.76667 1.56667 1.60000	1.90323 1.32258 1.41935 1.67742 1.54839	1.90000 1.63333 1.56667 1.70000 1.83333	1.90323 1.54839 1.74194 1.87097 1.58065
8 9 10 11 12	1982 1983 1984 1985 1986 1987	1.41935 2.00000 1.29032 1.58065 1.29032 1.74194	2.00000 1.54839 1.41935 1.45161 1.64516 1.48387	1.66667 1.83333 1.76667 1.56667 1.60000 1.56667	1.90323 1.32258 1.41935 1.67742 1.54839 1.77419	1.90000 1.63333 1.56667 1.70000 1.83333 1.66667	1.90323 1.54839 1.74194 1.87097 1.58065 2.32258
8 9 10 11 12 13	1982 1983 1984 1985 1986 1987 1988	1.41935 2.00000 1.29032 1.58065 1.29032 1.74194 1.51613	2.00000 1.54839 1.41935 1.45161 1.64516 1.48387 1.54839	1.66667 1.83333 1.76667 1.56667 1.60000 1.56667 1.90000	1.90323 1.32258 1.41935 1.67742 1.54839 1.77419 1.51613	1.90000 1.63333 1.56667 1.70000 1.83333 1.66667 1.26667	1.90323 1.54839 1.74194 1.87097 1.58065 2.32258 1.61290
8 9 10 11 12 13 14	1982 1983 1984 1985 1986 1987 1988 1988	1.41935 2.00000 1.29032 1.58065 1.29032 1.74194 1.51613 1.64516	2.00000 1.54839 1.41935 1.45161 1.64516 1.48387 1.54839 1.70968	1.66667 1.83333 1.76667 1.56667 1.60000 1.56667 1.90000 2.20000	1.90323 1.32258 1.41935 1.67742 1.54839 1.77419 1.51613 1.41935	1.90000 1.63333 1.56667 1.70000 1.83333 1.66667 1.26667 1.13333	1.90323 1.54839 1.74194 1.87097 1.58065 2.32258 1.61290 1.87097

.

TABLE A.6 New Motor Vehicle Registrations

Row	YEAR	JAN	FEB	MAR	APR	MAY	JUN
1	1981	*	*	11171	10385	11246	13443
2	1982	8862	9695	15345	11152	12868	17281
3	1983	8486	9698	13724	11426	13088	12254
4	1984	10032	11990	14058	12253	15789	13425
5	1985	11311	16209	17502	14391	16847	13588
6	1986	12182	11622	10710	14982	11466	13407
7	1987	10990	9535	11359	8726	10235	10713
8	1988	8456	9123	9615	10007	11739	10890
9	1989	8321	10850	13877	10474	11197	13761
10	1990	8440	12051	19521	14863	15066	12378
ROW	YEAR	JUL	AUG	SEP	ост	NOV	DEC
1	1981	18013	11960	12073	10068	9053	13107
2	1982	14529	14584	11984	14162	11765	14820
3	1983	11704	13236	12301	11903	13395	13216
4	1984	16171	14053	12650	15128	13193	11798
5	1985	15475	13751	14039	15766	13236	11925
6	1986	12603	13076	10682	11916	10135	8303
7	1987	11078	8640	11247	11151	7343	12872
8	1988	9237	13655	12012	11169	14441	12500
9	1989	16435	13875	17522	12090	13306	13818
10	1990	15627	12055	12701	13310	10543	8680

.

#### TABLE A.7 Weather Index

RO₩	YEAR	JAN	FEB	MAR	APR	MAY	NUL
1	1981	•	*	9.9461	7.9461	16.9538	15.9846
2	1982	6.9769	4.9769	9.9461	7.9769	12.9308	8.9846
3	1983	8.9461	1.0077	9.9615	12,9461	15.9231	14.9461
4	1984	12.9538	6.9615	9.9615	10.9538	5.9692	7.9384
5	1985	5.9769	4.9615	9.9308	9.9461	11.9461	14.9308
6	1986	6.9461	3.9769	3.9692	15.9000	16.9384	11.9384
7	1987	8.9692	6.9692	9.9461	7.9461	16.9154	14.9692
8	1988	5.9692	1.9923	9,9384	5.9923	15.9615	12.9538
9	1989	8.9461	5.9615	12.9231	10,9308	10.9692	15.9461
10	1990	3.0000	13.9154	6.9615	13.9231	9.9692	15.9308
ROW	YEAR	JUL	AUG	SEP	001	NOV	DEC
1	1981	14.9846	19.9384	7,0000	10.9461	12,9384	7.9615
2	1982	6.9769	8.9538	11.9846	9.9538	4.9692	5.9692
3	1983	14.9692	11.9692	19.9077	11.9615	14.9384	5.0154
4	1984	15,9615	19.9538	15.9308	10.9538	9.9615	13.9077
5	1985	12.9384	19.9923	13.9384	11.9615	15.9231	17.9231
6	1986	17.9538	12.9923	13.9461	20.9231	8.9615	9.9692
7	1987	11.0000	15.9461	13.9384	9.9769	12.9231	9.9538
8	1988	17.9461	15.9154	12.9615	16.8769	14.9384	11.9461
9	1989	17.9231	22.8846	12.9461	17.8923	9,9461	9.9615
10	1990	18.9384	23.9308	9.9538	15.9077	9.9538	7.9846

:

QUEENSLAND MONTHLY DATA TABLE A.8 Number of fatal road crashes by day

ROW	YEAR	JAN	FEB	MA	R A	PR	MAY	JUN
1	1976	0 77419	1 17241	1 2580	د 1 TTT	27 1	15/.8/ 1	22222
2	1077	1 22581	0 79571	1 4120	0 1.333		(1200 1	20000
27	1070	1 7000	1 1 70571	1.0127				.20000
3	1970	1.32230	1.1/07/	1.2903				.2000/
4	1979	0.90774	1.21429	1.2258	1 1.033	1.4	+5161 1	./000/
2	1980	1.00000	0.93103	1.3871	0 1.366	6/ 1.	8065 1	.90000
6	1981	1.12903	1.17857	1.5161	5 1.233	33 1,0	06452 1	.83333
7	1982	1.19355	1.28571	1.1935	5 1.533	i <b>33</b> 1.6	51290 1	.56667
8	1983	1_06452	1.03571	1.0000	0 1.000	00 0.9	96774 1	.36667
9	1984	0.90323	0.79310	1.2580	6 1.433	33 1.1	12903 0	.93333
10	1985	1.12903	1,25000	1.4838	7 1.100	00 1.2	22581 1	.40000
11	1986	1.16129	1.42857	1.2903	2 1.333	33 0.9	93548 1	.16667
12	1987	0.90323	1.03571	1.2258	1 0.766	67 0.9	70323 1	. 13333
13	1988	1.29032	1.35714	1.0645	2 1.066	67 1.4	41935 1	.06667
14	1989	0.80645	0.75000	1.0645	2 0.733	33 0.9	96774 0	.90000
15	1990	0.61290	0.85714	0.9032	3 0.866	67 0.1	77419 1	.16667
ROW	YEAR	JUL	AUG	SEi	P C	ст	NOV	DEC
1	1976	1.12903	1.16129	1.6000	0 1.645	16 1.6	63333 1	.90323
2	1977	1.58065	1.45161	1.3666	7 1.580	165 1.4	40000 1	.74194
3	1978	1.38710	1.83871	1.8000	0 1.903	23 1.4	66667 1	,70968
4	1979	1,58065	1.74194	1.6333	3 1.451	61 1.6	53333 1	.58065
5	1980	1.54839	1.35484	1.1666	7 1.516	13 1.6	63333 1	.25806
6	1981	1.38710	1.35484	1.7666	7 1.516	13 1.5	53333 1	.25806
7	1982	1.54839	1.48387	1.5000	0 1.451	61 1.0	53333 1	.16129
8	1983	0.83871	1.67742	1.4666	7 1.193	55 1.1	16667 1	.58065
9	1984	0.83871	1.51613	1.6666	7 1.483	87 1.3	30000 1	.41935
10	1985	1,00000	1.06452	1.4000	0 1.516	13 1.0	06667 1	,22581
11	1986	1,25806	1.06452	1.2333	3 0.903	23 1.0	3333 1	.06452
12	1987	1,22581	1.35484	1.2333	3 1.193	55 1.3	33333 0	.83871
13	1988	1,29032	1.67742	1.4666	7 1.387	10 1.0	53333 1	.16129
14	1989	1,12903	0.90323	1 1000	0 1.161	29 1.4	46667 1	35484
15	1990	0.96774	1-12903	1.2333	3 0.967	74 0.9	20000 1	12903
TABLE	A.9 We	ather Index						
ROW	YEAR	JAN	FEB	MAR	APR	MAY	JUN	
1	1981	*	*	7.602	10.384	7.641	2.951	
2	1982	14.277	15.379	16.384	11.772	10.282	2.131	
3	1983	9,049	12.635	13.296	11.559	18.971	11.470	1
4	1984	14.097	10.815	6.510	8.462	10.243	11.097	,
5	1985	9,111	16.296	18.399	9.160	15.034	3.136	
6	1986	8.821	10.097	10.171	7.097	13.020	4.550	1
7	1987	11.898	15.490	9.830	14.875	14.854	13.863	i
8	1988	17.801	7.744	15.641	20.616	4.613	8.655	
9	1989	12.311	12,199	15.325	16.399	20.020	8.835	
10	1990	9.057	13.490	16.296	16.986	11.641	9.932	
ROW	YEAR	JUL	AUG	SEP	OCT	NOV	DEC	
1	1981	8.359	2.869	4.684	10.160	15.932	14.533	i
2	1982	4.000	13.330	9.567	5.601	8,140	15.504	,
3	1983	9.752	8.155	4.117	9,767	14.951	10.481	
4	1984	11.878	5.684	8.476	15.305	11.023	10.257	,
5	1985	6.849	5.014	7.587	12.330	11.820	10.801	
6	1986	10,946	10.849	5.077	9.592	7.995	13.373	i
7	1987	7.786	9.767	2.519	11.339	8.408	11.057	•
8	1988	8,442	10.151	6.864	3.291	8.393	18.165	i
9	1989	7.088	4.738	4.732	7.567	13.869	10.655	i
10	1990	4.111	4.621	3.505	5.704	7.616	*	I

#### AUSTRALIA QUARTERLY DATA

TABLE A.10. Number of fatal road crashes standardised by estimated population ('000)

YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1981	*	0.048	0.050	0.051
1982	0.046	0.048	0.047	0.047
1983	0.036	0.040	0.043	0.041
1984	0.039	0.038	0.039	0.044
1985	0.039	0.041	0.039	0.047
1986	0.041	0.041	0.037	0.041
1987	0.036	0.036	0.038	0.042
1988	0.039	0.036	0.041	0.039
1989	0.036	0.034	0.036	0.036
1990	0.029	0.030	0.032	٠
	YEAR 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990	YEAR         MARCH Q.           1981         *           1982         0.046           1983         0.036           1984         0.039           1985         0.039           1986         0.041           1987         0.036           1988         0.039           1980         0.036	YEAR         MARCH Q.         JUNE Q.           1981         *         0.048           1982         0.046         0.048           1983         0.036         0.040           1984         0.039         0.038           1985         0.039         0.041           1986         0.041         0.041           1988         0.039         0.036           1989         0.036         0.034           1990         0.029         0.030	YEAR         MARCH Q.         JUNE Q.         SEPT Q.           1981         *         0.048         0.050           1982         0.046         0.048         0.047           1983         0.036         0.040         0.043           1984         0.039         0.038         0.039           1985         0.039         0.041         0.037           1986         0.041         0.041         0.037           1987         0.036         0.036         0.038           1988         0.039         0.036         0.041           1989         0.036         0.034         0.036           1990         0.029         0.030         0.032

TABLE A.11. Percentage change in average retail price of petrol over eight capital cities

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	0.101	-0.041	0.025
ż	1982	0.002	0.020	0.104	0.021
3	1983	0.037	-0.030	0.053	0.047
4	1984	-0.002	0.014	0.010	0.002
5	1985	0.020	0.085	0.034	-0.031
6	1986	0.015	-0.107	0.065	0.077
7	1987	0.018	-0.017	-0.001	0.020
8	1988	0.014	-0.046	-0.015	-0.019
9	1989	-0.002	0.083	0.014	0.032
10	1990	0.062	-0.004	0.045	0.206

TABLE A.12. Automotive fuel sales (megalitres)

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	3716171	3823972	3839872
2	1982	3699134	3861840	3720533	3904406
3	1983	3607371	3751118	3784060	3882418
4	1984	3802535	3867537	3801619	4014809
5	1985	3839181	3923407	*	4100403
6	1986	3815766	4064484	4012536	4110137
7	1987	3879669	4005934	4097943	*
8	1988	4140660	4097406	4217293	*
9	1989	*	4294497	4308409	4388843
10	1990	4261324	4303956	4392969	4165786

TABLE A.13. New motor vehicle registrations standardised by estimated population ('000)

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	0.010	0.011	0.010
2	1982	0.010	0.011	0.011	0.010
3	1983	0,009	0.009	0.010	0.009
4	1984	0.010	0.011	0.010	0.010
5	1985	0.011	0.011	0.011	0.010
6	1986	0.008	0.009	0.009	0.007
7	1987	0,007	0.007	0,007	0.007
8	1988	0.007	0.008	0,008	0.009
9	1989	0.008	0,009	0.010	0.009
10	1990	0.009	0.009	0.009	•

#### VICTORIA QUARTERLY DATA

### TABLE A.14. Number of fatal road crashes standardised by estimated population (\*000)

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	0.044	0.044	0.041
2	1982	0.036	0.040	0.039	0.044
3	1983	0,032	0.041	0.041	0.034
4	1984	0.039	0.033	0.034	0.035
5	1985	0.032	0.040	0.034	0.039
6	1986	0.038	0.038	0.033	0.036
7	1987	0.034	0.038	0.035	0.042
8	1988	0.038	0.039	0.036	0.031
9	1989	0.045	0.041	0.039	0.037
10	1990	0.030	0.032	0.025	*

#### TABLE A.15. New motor vehicle registrations standardised by estimated population ('000)

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	0.009	0.011	0.008
2	1982	0.009	0.010	0.010	0.010
3	1983	0.008	0.009	0.009	0.009
4	1984	0.009	0.010	0.010	0.010
5	1985	0.011	0.011	0.010	0.010
6	1986	0.008	0.010	0.009	0,007
7	1987	0.008	0.007	0.007	0.007
8	1988	0.006	0.008	0.008	0,009
9	1989	0.008	0.008	0.011	0.009
10	1990	0.009	0.010	0.009	*

### TABLE A.16. Percentage change in average retail price of petrol price over eight capital cities

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	0.092	-0.087	0.040
2	1982	-0.075	0.093	0.137	-0.036
3	1983	0,052	-0.052	0.118	0.043
4	1984	-0.017	0.042	0.002	-0.010
5	1985	0.023	0.126	0.029	-0.079
6	1986	0.079	-0.124	0.055	0.066
7	1987	0.033	-0.046	-0.007	0.063
8	1988	-0.037	-0.049	-0.015	-0.031
9	1989	0.032	0.072	0.005	0.043
10	1990	0.066	0.000	0.053	0.242

#### TABLE A.17. Automotive fuel sales (megalitres)

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	*	1034901	1055658	1094359
2	1982	1043062	1079788	1059513	1097663
3	1983	1014437	1073603	1054540	1079659
4	1984	1071671	1081560	1056898	1129183
5	1985	1080998	1106391	*	1159275
6	1986	1066838	1136422	1099522	1171431
7	1987	1080793	1130201	1141658	±
8	1988	1152917	1167708	1146720	×
9	1989	٠	1221678	1193184	1236323
10	1990	1190300	1197615	1191160	1157554

TABLE A.18. Automotive diesel oil sales (megalitres)

;

ROW	YEAR	MARCH Q.	JUNE Q.	SEPT Q.	DEC Q.
1	1981	•	328063	307713	303089
2	1982	291014	343115	320000	297579
3	1983	281661	351201	319165	325362
4	1984	317400	357860	320083	347083
5	1985	326051	370044	*	349882
6	1986	315554	377508	347841	359668
7	1987	345781	383533	360290	*
8	1988	370726	395168	384360	*
9	1989	*	446903	405100	443322
10	1990	429172	439017	399832	372875

Time Series Plot of Standardised Monthly Fatal Road Accidents for Australia



Figure A.2

Time Series Plot of Standardised Monthly Fatal Road Accidents for NSW







Figure A.4

#### Time Series Plot of Standardised Monthly Fatal Road Accidents for Victoria



Figure A.5



Figure A.6



Figure A.7

į

ļ

#### Fatal Road Accidents for Australia



Months (beginning January 1976)

Monthly Fatal Road Accidents(std by day)

Time Series Plot of Standardised Monthly Fatal Road Accidents for Queensland



Figure A.B.





Figure A.9.





Quarters (beginning June quarter 1981)

Figure A.10

Time Series Plot of Percentage Change in Petrol Price for Australia



Figure A.11

Time Series Plot of Automotive Fuel Sales for Australia



Figure A.13



Figure A.15

Time Series Plot of Percentage Change in Petrol Prices for Victoria







Figure A.17

(× 100000)





Figure A.18

### **APPENDIX B**

### Sources of Explanatory Variables

.

.

Sources of Data and Definitions of Derived Variables

#### **Response Variables**

- 1. FATACC Number of Fatal Accidents
  - SOURCE: (i) ABS Publication No. 9401.0 'Road Traffic Accidents Involving Fatalities, Australia ' Monthly; first issue containing Table 2 - Fatal Road Traffic Accidents - July 1976; last issue December 1990.
    - (ii) FORS Publication
       'Road Crash Statistics,Australia'
       Monthly; first issue January 1991.
- DESCRIPTION: The number of fatal road traffic accidents in each State and Territory.
- 2. FATALS Number of Road Traffic Accident Fatalities
  - SOURCE: (i) ABS Publication No. 9401.0 Monthly; first issue January 1970; last issue December 1990.
    - (ii) FORS Publication
       'Road Crash Statistics,Australia'
       Monthly; first issue January 1991.
- DESCRIPTION: The number of persons killed in road traffic accidents.

#### Explanatory Variables

- 1. CPI Consumer Price Index for all Groups: six state capital cities and Canberra.
  - SOURCE: ABS publication No. 6401.0 'Consumer Price Index' Quarterly: first issue June 1960.
- CPITRS Consumer Price Index for Transportation Group: weighted average of six state capital cities.
  - SOURCE: ABS publication No. 6401.0 Data available for base year 1980-81 from March quarter 1982.
- 3. GDP Gross Domestic Product
  - SOURCE: ABS Publication No. 5206.0 'Australian National Accounts: National Income and Expenditure' December Quarter 1990; Table 42

DESCRIPTION: Gross Domestic Product for Australia at current

prices (\$million). Data collected from September quarter 1975.

4. NEWMVR Registrations of New Motor Vehicles

SOURCE: ABS Publication No. 9301.0 'Registrations of New Motor Vehicles, Australia, Preliminary' Monthly(Australia);first issue September 1953. State and Territory data available quarterly to May 1979 then monthly from June 1979. Data collected from March quarter 1975.

- DESCRIPTION: Registrations in each State and Territory of new motor vehicles.
- 5. PETROL Average Retail Price of Petrol (super grade)

- DESCRIPTION: Average retail prices of super grade petrol for each of the six State capitals, Canberra and Darwin. Prices are collected at the middle of the first month of the quarter.
- 6. POPN Resident Population Estimates
  - SOURCE: ABS Publication No. 3101.0 'Australian Demographic Statistics' Quarterly: first issue June 1979. Data collected from June guarter 1971.
- DESCRIPTION: Latest quarterly population estimates('000) for Australia, States and territories.
  - POPNAGE Annual Estimated Resident Population by Age Groups, States and Territories.
  - POPNFEM} Annual Estimated Resident Population by age and POPNMAL} sex groups, States and Territories.
  - SOURCE: ABS Publication No: 3201.0 'Estimated Resident Population by Sex and Age: States and Territories of Australia' Annual; first issue 30 June 1968.
- 7. RDAYS Number of Rain Days
  - SOURCE: Australian Bureau of Meteorology 'Report of Monthly and Yearly Rainfall by N.C.C.' for chosen stations within each State and

Territory. Data was collected from 1970 onwards.

DESCRIPTION: Number of days for which rainfall is recorded at a chosen station within a given month. The 17 stations chosen were:

> NSW: Sydney, Dubbo, Newcastle ACT: Canberra QLD: Brisbane, Cairns, Mackay SA : Adelaide, Port Augusta WA : Perth, Albany VIC: Melbourne, Mildura NT : Darwin, Alice Springs TAS: Hobart, Launceston

- 8. SFUEL Sales of Automotive Gasoline by State Marketing Area (kilolitres)
  - SDIESEL Sales of Inland Automotive Diesel Oil by State Marketing Area (kilolitres)
  - SLPG Sales of LPG for Automotive Use by State Marketing Area (kilolitres)
  - SOURCE: (i) Department of Primary Industries and Energy Bulletin 'Major Energy Statistics':ISSN 0727-260X Table 3B. Monthly Bulletins Available from March 1981 to May 1989.
    - (ii) Australian Bureau of Agricultural and Resource Economics (ABARE) Bulletin
       'Quarterly Mineral Statistics'
       First issue June quarter 1989.
       Monthly data from June 1989 was provided by
       ABARE on request.
- DESCRIPTION: Sales of petroleum products by States and Territories (kilolitres).
- 9. UNEMP Unemployment Rate

SOURCE: ABS Publication No: 6203.0 'The Labour Force, Australia' Monthly; data available monthly for states from from 1978.

DESCRIPTION: Unemployment Rate (expressed as a percentage) for the civilian population aged 15 and over for Australia, States and Territories.

#### ADDITIONAL VARIABLES INVESTIGATED

Data for additional variables that were sought but not readily available in the form required consists of:

1. The number of fatal road crashes and fatalities broken down into regions within states.

ABS publications available for regions within States are as follows:

- (i) VICTORIA 9406.2 'Road Traffic Accidents Involving Casualties, VIC'; annual.
- (ii) QUEENSLAND 9405.3 'Road Traffic Accidents, QLD'; quarterly 9406.3 'Road Traffic Accidents, QLD'; annual
- (iii) WESTERN AUSTRALIA
  - 9405.5 'Road Traffic Accidents involving Casualties Reported to the Police Department, WA'; quarterly
  - 9406.5 'Road Traffic Accidents Involving Casualties Reported to the Police Department, WA'; annual
- (iv) TASMANIA
  9405.6 'Road Traffic Accidents Involving Casualties,
  TAS'; quarterly
  9406.6 'Road Traffic Accidents Involving Casualties,
  TAS'; annual

Quarterly data is provided for Queensland, Tasmania and W.A. while monthly data is not published for any of the States or Territories.

2. The number of driver licenses in force.

This data could not be readily obtained for any of the states except Western Australia for which quarterly data is published in the publication no. 9406.5.

3. The number of vehicle kilometres travelled (vkt).

This data could only be located in the ABS Publication

9208.0 'Survey of Motor Vehicle Use, Australia' three-yearly; first issue September 1971; last issue September 1988.

The relevant information is contained in Table 15 'Total kilometres travelled by type of vehicle: State/Territory of registration twelve months ended 30 September 1988'. However as only 6 data values are available for vkt over 20 years this variable could not be included in our models containing monthly and quarterly data. It is suggested that the required data be collected monthly by some means other than survey, if possible, in the future.

4. The level of alcohol consumption.

As alcohol may be a cause of fatal road crashes, data on alcohol consumption was sought.

Annual data for alcohol consumption are shown in the ABS publication no. 4315.0 'Apparent consumption of selected foodstuffs, Australia, Preliminary';annual; first issue 1978-79. In the latest 1989-90 issue, a time series plot of 'Apparent per capita consumption of selected beverages 1984-85 to 1989-90, Australia, year ended 30 June' on page 3, shows that the consumption of wine has been steady while consumption of low alcohol beer has increased since 1988 and decreased steadily for other beer. The overall consumption of beer would appear to have remained fairly constant.

The annual data was not appropriate for inclusion in our models.

There is an ABS publication due for release in 1991 which contains national statistics describing the levels and patterns of alcohol consumption and selected demographic and socio-economic characterisics of consumers.

4381.0 'National Health Survey: Alcohol Consumption'

#### VARIABLES DERIVED FROM RAW DATA COLLECTED

For monthly data variables were standardised as follows:

1. STDACC

- (i) The number of monthly fatal accidents was standardised by the number of days in the month. This was done to eliminate the bias due to the length of the month.
- (ii) The variable was then standardised by population size to eliminate the effect of increasing population size over time.

So, STDACC = (STANDARDISED MONTHLY FATACC )/POPN

- %CHGPET This variable is a measure of the percentage change in petrol price from one month to the next.
- 3. MVR NEWMVR is standardised by dividing by population size.

so,

MVR = NEWMVR/POPN

4. WEATHER INDEX (WI)

The weather index is calculated by weighting the number of raindays for chosen centres within a State or Territory by their respective population size (Maunder, 1974). Population sizes for the centres chosen were taken from the ABS 1986 CENSUS 'Persons and Dwellings in legal local Government Areas, S.L.A.'s and Urban Centres/ Rural Localities'. Publication numbers 2462.0 - 2469.0.

The State WI is given by

$$RDS = \sum_{\substack{c \in S} \\ s \in S} (RDc \times Pc)$$
$$\sum_{\substack{c \in S \\ s \in S}} Pc$$

where

c denotes urban centre in state s denotes state or territory Pc denotes population size of centre RDc denotes number of rain days in centre

### APPENDIX C

### Computer Code for Predictive Models

.

-

```
41
         AUSTRALIA DATA
Regress standardised accidents on FUEL SALES,
 trend and quarters where fuel sales are produced in the
 transfer function model by a SARIMA(0,1,1,1,0,0,4) model.
41
UNITS[ NVALUES=39]
OPEN 'AUSQRT.DAT'; CHANNEL=2
OPEN 'FORFL91.DAT'; CHANNEL=3
READ[ CHANNEL=2] POPN, STDACC, UNEMP, NEWMVR, CHGPET, FUEL, DIESEL, TREND, \
01,02,03,04
READ [CHANNEL=3; SETNVALUES=Y] INTREND, INQ1, INQ2, INQ3, INQ4, INFUEL
VARIATE [NVALUES=4] FF
VARIATE DIFFACC, RES1, RES2, RES3, RES4, RES5
FACTOR [NVALUES=39; LEVELS=1(2,3,4,1)] QUART
GENERATE 4, QUART
CALC DIFFACC=DIFFERENCE(STDACC)
TREATMENT QUART
ANOVA [PRINT=A] DIFFACC
AKEEP QUART; VARIANCE=RVAR
CALC DSSS=RVAR*33
PRINT DSSS
CALC N=NOBS(STDACC)
CALC TSS=VAR(STDACC)*(N-1)
.
Using a SARIMA(1,0,1,1,1,0,0,4) model for errors
TSM ERM2; ORDERS=!(0,1,1)
TSM FU_ARIM; ORDERS=!(0,1,1,1,0,0,4)
ESTIMATE [PRINT=ESTIMATES] FUEL; TSH=FU_ARIM
FORECAST [MAXLEAD=4; FORECAST=FF]
TSM [model=1] FU_TSM; ORDERS=!(1,0,1,1,1,0,0,4)
TRANSFER TREND,Q2,Q3,Q4,FUEL; transfer=*,*,*,*,
ESTIMATE [PRINT=ESTIMATES; constant=f; MAXCYCLE=50] STDACC; TSM=ERM2
FORECAST [MAXLEAD=4] INTREND, INQ2, INQ3, INQ4, FF
TKEEP RESIDUALS=RES2
CALC N2=NOBS(RES2)
CALC ESS2=VAR(RES2)*(N2-1)
PRINT ESS2
CALC R22=(TSS-ESS2)/TSS
CALC R2S2=(DSSS-ESS2)/DSSS
```

```
PRINT R22,R2S2
```

```
AUSTRALIA DATA
Regress monthly standardised accidents on FUEL, LAG FUEL, LAG(LAGFUEL),
 trend and months, fitting structural models 1 to 5 as defined
below. Models fitted to time series from March 1981 to December
1990 and fatal accidents predicted for months January 1991 to
December 1992. Fuel sales predicted from airline model.
PROGRAM MLAG2.GEN
UNITS[ NVALUES=118]
OPEN 'MTHFL.DAT'; CHANNEL=2
OPEN 'MV.DAT' ; CHANNEL=3
READ [ CHANNEL=2] STDACC, FUEL, LAGFUEL, TREND, \
M1, M2, M3, H4, H5, M6, H7, H8, H9, H10, H11, H12
READ [CHANNEL=3; SETNVALUES=Y] INTREND. INM1. INM2. INM3. INM4. INM5. \
INM6, INM7, INM8, INM9, INM10, INM11, INM12
VARIATE DIFFACC, RES1, RES2, RES3, RES4, RES5, LLFUEL
VARIATE [NVALUES=24] FF,LAGFF,MVFF,LLAGFF,LLMVFF
VARIATE [NVALUES=24; VALUES=1386475,23(1)] REP
VARIATE [NVALUES=24; VALUES=1406254,1386475,22(1)] LREP
FACTOR [NVALUES=118; LEVELS=!(3,4,5,6,7,8,9,10,11,12,1,2)] MONTH
GENERATE 12, MONTH
CALC LLFUEL=SHIFT(LAGFUEL;1)
Remove trend by differencing and seasons by ANOVA
DSS = deseasonalised sums of squares
CALC DIFFACC=DIFFERENCE(STDACC)
TREATMENT MONTH
ANOVA [PRINT=A] DIFFACC
AKEEP MONTH; VARIANCE=RVAR
CALC DSSS=RVAR*105
PRINT DSSS
CALC N=NOBS(STDACC)
CALC TSS=VAR(STDACC)*(N-1)
Predicting fuel sales for 1991 and 1992 from past data,
using a SARIMA(0,1,1,0,1,1,12) model for errors.
TSM FU_ARIN; ORDERS=!(0,1,1,0,1,1,12)
ESTIMATE [PRINT=ESTIMATES] FUEL; TSM=FU_ARIM
FORECAST [MAXLEAD=24; FORECAST=FF]
CALC MVFF=SHIFT(FF;1)
CALC LAGFF=MVREPLACE(MVFF;REP)
CALC LLMVFF=SHIFT(FF;2)
CALC LLAGFF=MVREPLACE(LLMVFF;LREP)
PRINT LAGFF, LLAGFF
Using an MA(1) model for errors
TSM ERM1; ORDERS=!(0,0,1)
TRANSFER TREND, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, FUEL, LAGFUEL, LLFUEL
ESTIMATE [PRINT =ESTIMATES] STDACC; TSM=ERM1
FORECAST [MAXLEAD=24] INTREND, INM2, INM3, INM4, INM5, \
INM6, INM7, INM8, INM9, INM10, INM11, INM12, FF, LAGFF, LLAGFF
TKEEP RESID=RES1
CALC N1=NOBS(RES1)
CALC ESS1=VAR(RES1)*(N1-1)
PRINT ESS1
CALC R21=(TSS-ESS1)/TSS
CALC R2S1=(DSSS-ESS1)/DSSS
```
```
PRINT R21,R2S1
Using an IMA(0,1,1) model for errors
64
TSM ERM2; ORDERS=!(0,1,1)
TRANSFER TREND, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, FUEL, LAGFUEL, LL FUEL
ESTIMATE [PRINT=ESTIMATES; CONSTANT=F] STDACC; TSM=ERM2
FORECAST [MAXLEAD=24] INTREND, INM2, INM3, INM4, INM5, \
INM6, INM7, INM8, INM9, INM10, INM11, INM12, FF, LAGFF, LLAGFF
TKEEP RESID=RES2
CALC N2=NOBS(RES2)
CALC ESS2=VAR(RES2)*(N2-1)
PRINT ESS2
CALC R22=(TSS-ESS2)/TSS
CALC R2S2=(DSSS-ESS2)/DSSS
PRINT R22,R2SZ
Using an IMA(0,2,2) model for errors
TSM ERM3; ORDERS=!(0,2,2)
TRANSFER M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12, FUEL, LAGFUEL, LL FUEL
ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; ISM=ERM3
FORECAST [MAXLEAD=24] INM2,INM3,INM4,INM5,\
INM6, INM7, INM8, INM9, INM10, INM11, INM12, FF, LAGFF, LLAGFF
TKEEP RESID=RES3
CALC N3=NOBS(RES3)
CALC ESS3=VAR(RES3)*(N3-1)
PRINT ESS3
CALC R23=(TSS-ESS3)/TSS
CALC R2S3=(DSSS+ESS3)/DSSS
PRINT R23,R2S3
...
Using an ARIMA(0,2,2,0,1,1,12) model for errors
TSM ERM4; ORDERS= (0,2,2,0,1,1,12)
TRANSFER FUEL, LAGFUEL, LLFUEL
ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM4
FORECAST [MAXLEAD=24] FF, LAGFF, LLAGFF
TKEEP RESID=RES4
CALC N4=NOBS(RES4)
CALC ESS4=VAR(RES4)*(N4-1)
PRINT ESS4
CALC R24=(TSS-ESS4)/TSS
CALC R2S4=(DSSS-ESS4)/DSSS
PRINT R24,R2S4
н
using an ARIMA(0,1,1,0,1,1,12) model for errors
TSM ERM5; ORDERS=!(0,1,1,0,1,1,12)
TRANSFER FUEL, LAGFUEL, LLFUEL
ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM5
FORECAST [MAXLEAD=24] FF, LAGFF, LLAGFF
TKEEP RESID=RES5
CALC N5=NOBS(RES5)
CALC ESS5=VAR(RES5)*(N5-1)
PRINT ESS5
CALC R25=(TSS-ESS5)/TSS
CALC R2S5=(DSSS-ESS5)/DSSS
PRINT R25,R2S5
```

```
STOP
```

16-0CT-1991 16:16:43.00 Genstat 5 Release 2.1 (Vax/VMS5) Copyright 1990, Lawes Agricultural Trust (Rothamsted Experimental Station) 1 2 .... AUSTRALIA DATA -3 Regress standardised accidents on FUEL SALES, trend and quarters where fuel sales are produced in the -4 -5 transfer function model by a SARIMA(0,1,1,1,0,0,4) model. 0 -6 7 UNITS[ NVALUES=39] 8 OPEN 'AUSGRT.DAT'; CHANNEL=2 9 OPEN 'FORFL91.DAT'; CHANNEL=3 10 READ[ CHANNEL=2] POPN, STDACC, UNEMP, NEWMYR, CHGPET, FUEL, DIESEL, TREND, \ 11 01,02,03,04 Missing Identifier Minimum Mean Maximum Values POPN 14927 15985 17149 39 -1 0.02900 STDACC 0.04008 0.05100 39 1 10.400 30 UNEMP 5.467 7.709 n NEWMVR 0.007000 0.009211 0.011000 39 1 CHGPET -0.10700 0.02336 0.20600 39 0 3607371 3985657 4392969 39 FUEL 4 1658003 2109155 2612561 39 DIESEL 4 20.00 39.00 39 1.00 n TREND 39 0.0000 0.2308 01 1.0000 Ω Skew ۹2 0.0000 0.2564 1.0000 39 0 ۵3 0,0000 0.2564 1.0000 39 0 **Q**4 0.0000 0.2564 1.0000 39 0 12 READ [CHANNEL=3; SETNVALUES=Y] INTREND, INQ1, INQ2, INQ3, INQ4, INFUEL Identifier Minimum Mean Maximum Values Missing INTREND 40.00 41.50 43.00 4 0 0.0000 1.0000 4 INQ1 0.2500 0 0.0000 INQ2 0.2500 1.0000 4 ۵ 1.0000 0.0000 0.2500 INQ3 4 ۵ INQ4 0.0000 0.2500 1.0000 4 0 4202420 4270318 4336290 4 0 INFUEL 13 VARIATE [NVALUES=4] FF 14 VARIATE DIFFACC, RES1, RES2, RES3, RES4, RES5 15 FACTOR [NVALUES=39; LEVELS=!(2,3,4,1)] QUART 16 GENERATE 4, QUART \*\*\*\*\*\*\* Warning (Code CA 24). Statement 1 on Line 16 Command: GENERATE 4,QUART

Number of units of GENERATED factors does not give an exact number of reps

17 CALC DIFFACC=DIFFERENCE(STDACC)

- 18 TREATMENT QUART
- 19 ANOVA [PRINT=A] DIFFACC

19..... \*\*\*\*\* Analysis of variance \*\*\*\*\* Variate: DIFFACC Source of variation d.f.(m.v.) /.) s.s. m.s. 0.292E-03 0.973E-04 0.252E-03 0.762E-05 ν.Γ 3 33(2) QUART 12.76 Residual Total 36(2) 0.537E-03 20 AKEEP QUART; VARIANCE=RVAR 21 CALC DSSS=RVAR\*33 22 PRINT DSSS DSSS 0.0002516 23 24 CALC N=NOBS(STDACC) 25 CALC TSS=VAR(STDACC)\*(N-1) 26 27 - 0 -28 Using a SARIMA(1,0,1,1,1,0,0,4) model for errors -29 u 30 TSM ERM2; ORDERS=!(0,1,1) 31 32 TSM FU\_ARIM; ORDERS=!(0,1,1,1,0,0,4) 33 ESTIMATE [PRINT=ESTIMATES] FUEL; TSM=FU\_ARIM

---

\*\*\*\*\* Time-series analysis \*\*\*\*\*

\*\*\* Autoregressive moving-average model \*\*\*

Innovation variance 8.112E+09

Transformation Constant	ref. 0 1	estimate 1.00000 15842.	s.e. Fixed 2276.
• Non-seasonal;	differencin	ng order 1	
Moving-average	lag ref. 1 2	estimate 0,994	s.e. 0.105
* Seasonal; per	iod 4; no di	fferencing	
Autoregressive	lag ref. 4 3	estimate 0.495	s.e. 0.197

34 FORECAST [MAXLEAD=4; FORECAST=FF]

34....

\*\*\* Forecasts \*\*\*

Maximum lead time: 4

Lead	time	forecast	lower limit	upper limit
	1	4323477.	4175333.	4471621
	2	4352584.	4204438,	4500731.
	3	4404657.	4256508.	4552806.
	4	4300163.	4152012.	4448315.

٠

.

.

35 36 TSM FU\_TSM; ORDERS=!(1,0,1,1,1,0,0,4); PARAMETERS=!(1,0,0,0.1) 37 TRANSFER TREND,02,03,04,FUEL; ARIMA=\*,\*,\*,FU\_TSM 38 ESTIMATE[PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM2

38			
**** Time-seri	es analysis	****	
*** Transfer-fu	nction mode	l 1 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 1	estimate -0.000812	s.e. 0.000308
*** Transfer-fu	nction mode	l 2 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 2	estimate -0.000624	s.e. 0.000807
*** Transfer-fu	nction mode	l 3 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
• Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 3	estimate 0.001238	s.e. 0.000891
*** Transfer-fu	nction mode	( 4 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 4	estimate 0.00149	s.e. 0.00124
*** Transfer-fu	nction mode	l 5 ***	

.

•

- - - - - - -

1

- ----

Delay time O

-

Transformation	ref.	estimate	S.e.
Constant	õ	0.	FIXED
* Non-seasonal;	no differer	ncing	
Moving-average	lag ref. 0 5	estimate 0.18E-07	s.e. 0.52E-08
*** Autoregressi	ve moving-a	overage model	***
Innovation varia	nce 0.00000	4417	
_ / .	ref.	estimate	s.e.
Transformation Constant	0	1.00000 0.	FIXED FIXED
* Non-seasonal; (	differencin	g order 1	
	lag ref.	estimate	s.e.
moving-average	ID	0.154	U.199
39			

40 FORECAST [MAXLEAD=4] INTREND, INQ2, INQ3, INQ4, FF; METHOD=0,0,0,0,F

-

.

40..... \*\*\* Forecasts \*\*\* Maximum lead time: 4 Lead time forecast lower limit upper limit 1 0.02794 0.02448 0.03140 0.02251 0.02704 2 0.03156 0.02904 0.02657 3 4 0.03443 0.02044 0.03270 41 41
42 TKEEP RESIDUALS=RES2
43 CALC N2=NOBS(RES2)
44 CALC ESS2=VAR(RES2)\*(N2-1)
45 PRINT ESS2 ESSZ 0.0001193 46 47 CALC R22=(TSS-ESS2)/TSS 48 CALC R2S2=(DSSS-ESS2)/DSSS 49 PRINT R22, R2S2 R22 R2S2 0.8834 0,5259 50 STOP

> i i

> > ÷

\*\*\*\*\*\*\*\* End of job. Maximum of 19152 data units used at line 43 (30562 left)

Genstat 5 Release 2.1 (Vax/VMS5) 30-007-1991 09:21:35.62 Copyright 1990, Lawes Agricultural Trust (Rothamsted Experimental Station) 1 2 N AUSTRALIA DATA -3 Regress monthly standardised accidents on FUEL, LAG FUEL, LAG(LAGFUEL), trend and months, fitting structural models 1 to 5 as defined below. Models fitted to time series from March 1981 to December -4 -5 1990 and fatal accidents predicted for months January 1991 to -6 -7 December 1992. Fuel sales predicted from airline model. -8 -9 PROGRAM MLAG2.GEN - 10 UNITS[ NVALUES=118] 11 OPEN 'MTHFL.DAT'; CHANNEL=2 OPEN 'MV.DAT' ; CHANNEL=3 READ[ CHANNEL=2] STDACC,FUEL,LAGFUEL,TREND,\ 12 13 OPEN 'NV.DAT' 14 15 M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12 Identifier Minimum Mean Maximum Values Missing STDACC 4.774 6.960 8.968 118 0 FUEL 1118114 1330823 1639366 118 4 1118114 1639366 LAGFUEL 1330331 118 5 TREND 1.00 59.50 118.00 118 ٥ 0.00000 0.07627 M1 1.00000 118 Û Skew 0.00000 Μ2 0.07627 1.00000 118 Skew ۵ M3 0.00000 0.08475 1.00000 118 Skew Û M4 0.00000 0.08475 1.00000 118 0 Skew М5 0,00000 0.08475 1.00000 118 Ô. Skew 0.00000 0.08475 1.00000 M6 118 Ω Skew M7 0.00000 1.00000 0.08475 118 Û Skew Μ8 0.00000 0.08475 1.00000 118 0 Skew Μ9 0,00000 0.08475 1.00000 118 0 Skew M10 0,00000 0.08475 1.00000 118 0 Skew 0.00000 0.08475 1 00000 M11 118 D Skeu 0.00000 M12 0.08475 1.00000 118 ۵ Skew 16 READ [CHANNEL=3; SETNVALUES=Y] INTREND, INM1, INM2, INM3, INM4, INM5, \ 17 INM6, INM7, INM8, INM9, INM10, INM11, INM12 Identifier Missing Minimum Maximum Values Mean INTREND 119.0 130.5 142 0 24 0 TNM1 0.00000 0.08333 1.00000 24 0 Skew 1NM2 0,00000 0.08333 1.00000 24 0 Ske⊯ INM3 0,00000 0.08333 1.00000 24 0 Skew 0.00000 1NM4 0.08333 1.00000 Skew 24 D 0.00000 INM5 0.08333 1.00000 24 ΰ Skev INM6 0,00000 0.08333 1.00000 24 0 Skew INM7 0.00000 0.08333 1.00000 24 Û Skew INM8 0.00000 0.08333 1.00000 24 ٥ Skew 1.00000 INM9 0.00000 0.08333 24 Skew Û 0.00000 1.00000 24 INM10 0.08333 n Skew INM11 0.00000 0.08333 1.00000 24 D Skew 0.00000 0.08333 1.00000 INM12 24 0 Skew 18 VARIATE DIFFACC, RES1, RES2, RES3, RES4, RES5, LLFUEL VARIATE INVALUES=241 FF, LAGFF, MVFF, LLAGFF, LLMVFF 19 20 VARIATE [NVALUES=24; VALUES=1386475,23(1)] REP 21 VARIATE [NVALUES=24; VALUES=1406254,1386475,22(1)] LREP 22 FACTOR [NVALUES=118; LEVELS=!(3,4,5,6,7,8,9,10,11,12,1,2)] MONTH 23 GENERATE 12, MONTH

\*\*\*\*\*\*\* Warning (Code CA 24). Statement 1 on Line 23 Command: GENERATE 12,MONTH

Number of units of GENERATED factors does not give an exact number of reps

24 25 CALC LLFUEL=SHIFT(LAGFUEL;1) 26 27 " -28 Remove trend by differencing and seasons by ANOVA -29 DSS = deseasonalised sums of squares -30 " 31 CALC DIFFACC=DIFFERENCE(STDACC) 32 TREATMENT MONTH 33 ANOVA [PRINT=A] DIFFACC

\_

33..... \*\*\*\*\* Analysis of variance \*\*\*\*\* Variate: DIFFACC Source of variation d.f.(m.v.) s.s. m.s. v.r. 28.4730 63.1615 91.4123 11 105(1) 2.5885 MONTH 4.30 Residual 0.6015 Total 116(1) 34 AKEEP MONTH; VARIANCE=RVAR
 35 CALC DSSS=RVAR\*105
 36 PRINT DSSS DSSS 63.16 37 38 CALC N=NOBS(STDACC) 39 CALC TSS=VAR(STDACC)\*(N-1) 40 4 -41 Predicting fuel sales for 1991 and 1992 from past data,
-42 using a SARIMA(0,1,1,0,1,1,12) model for errors.
-43 " 44 TSM FU\_ARIM; ORDERS=!(0,1,1,0,1,1,12) 45 ESTIMATE [PRINT=ESTIMATES] FUEL; TSM=FU\_ARIM

\_

45.....

----

\*\*\*\*\*\*\*\* Warning (Code TS 21). Statement 1 on Line 45 Command: ESTIMATE [PRINT=ESTIMATES] FUEL; TSM=FU\_ARIM

The iterative estimation process has not converged The maximum number of cycles is 15

\*\*\*\*\* Time-series analysis \*\*\*\*\*

\*\*\* Autoregressive moving-average model \*\*\*

Innovation variance 4.097E+09

	ref.	estimate	s.e.
Transformation	0	1.00000	FIXED
Constant	1	154.	178.

\* Non-seasonal; differencing order 1

	lag r	ef.	estimate	s.e.
Moving-average	1	2	0.9988	0.0564

Seasonal; period 12; differencing order 1

	lag	ref.	estimate	s.e.
Moving-average	12	3	0.650	0.116

46 FORECAST [MAXLEAD=24; FORECAST=FF]

46....

\*\*\* Forecasts \*\*\*

Maximum lead time: 24

Lead time	forecast	lower limit	upper limit
1	1457263.	1351977.	1562549.
2	1373949.	1268663.	1479235.
3	1507727.	1402440.	1613013.
4	1426119.	1320833.	1531405.
5	1516573.	1411287.	1621860.
6	1468163.	1362877.	1573450.
7	1482711.	1377424.	1587997.
8	1563090.	1457804.	1668377.
9	1424702.	1319415.	1529988.
10	1472606.	1367319.	1577892.
11	1481660.	1376373.	1586947.
12	1499614.	1394327.	1604901.
13	1489660.	1378075.	1601246.
14	1406502.	1294916.	1518087.
15	1540433.	1428847.	1652019.
16	1458980.	1347394.	1570566.
17	1549589.	1438003.	1661175.
18	1501334.	1389747.	1612920.
19	1516035.	1404449.	1627622.
20	1596570.	1484983.	1708156,
21	1458335.	1346749.	1569922.
22	1506394.	1394807.	1617981.
23	1515603.	1404016	1627190.
24	1533711.	1422124.	1645298.

- 47 CALC MVFF=SHIFT(FF;1)
  48 CALC LAGFF=MVREPLACE(MVFF;REP)
  49 CALC LLMVFF=SHIFT(FF;2)
  50 CALC LLAGFF=MVREPLACE(LLMVFF;LREP)
  51 PRINT LAGFF,LLAGFF

LAGFF	LLAGFF
1386475	1406254
1457263	1386475
1373949	1457263
1507727	1373949
1426119	1507727
1516573	1426119
1468163	1516573
1482711	1468163
1563090	1482711
1424702	1563090
1472606	1424702
1481660	1472606
1499614	1481660
1489660	1499614
1406502	1489660
1540433	1406502
1458980	1540433
1549589	1458980
1501334	1549589
1516035	1501334
1596570	1516035
1458335	1596570
1506394	1458335

#### 

- 52 " -53 Using an MA(1) model for errors -54 " 55 TSM ERM1; ORDERS=!(0,0,1) 56 TRANSFER TREND,M2,M3,M4,M5,M6,M7,M8,M9,M10,M11,M12,FUEL,LAGFUEL,LLFUEL 57 ESTIMATE [PRINT =ESTIMATES] STDACC; TSM=ERM1

57				
***** Time-seri	es analysis	****		
*** Transfer-fu	nction mode	l 1 ***		
Delay time O				
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED	
* Non-seasonal;	no differe	ncing		
Moving-average	lag ref. 0 1	estimate -0.02700	s.e. 0.00577	
*** Transfer-fu	nction mode	l 2 ***		
Delay time O				
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED	
* Non-seasonal;	no differe	ncing		
Moving-average	lag ref. O 2	estimate 0.766	s.e. 0.335	
*** Transfer-fu	nction mode	L 3 ***		
Delay time O				
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED	
* Non-seasonal;	no differe	ncing		
Moving-average	lag ref. 0 3	estimate 1.351	s.e. 0.391	
*** Transfer-function model 4 ***				
Delay time O				
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED	
* Non-seasonal;	no differe	n¢ing		
Moving-average	lag ref. 0 4	estimate 0.557	s.e. 0 <b>.3</b> 56	

.

\*\*\* Transfer-function model 5 \*\*\*

Delay time O

Terroformation	ref.	estimate	S.e.
Constant	0	0.	FIXED
* Non-seasonal;	no differe	ncing	
Noving-average	lag ref. 0 5	estimate 0.848	s.e. 0.325
*** Transfer-fu	nction mode	l 6 ***	
Delay time O			
Transformation Constant	ref. Q O	estimate 1.00000 0.	s.e. Fixed Fixed
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 6	estimate 0.782	s.e. 0.318
*** Transfer-fu	nction mode	( 7 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 7	estimate 0.733	s.e. 0.322
*** Transfer-fu	nction mode	L 8 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	S.e. FIXED FIXED
• Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 8	estimate 0.650	s.e. 0.326
*** Transfer-fu	nction mode	l 9 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. F1XED F1XED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 9	estimate 1.459	s.e. 0.324

-----

### \*\*\* Transfer-function model 10 \*\*\*

Delay time 0

Transformation Constant	ref. O O	estimate 1.00000 D.	S.e. Fixed Fixed				
* Non-seasonal;	* Non-seasonal; no differencing						
Moving-average	lag ref. O 10	estimate 1.140	s.e. 0.337				
*** Transfer-fu	nction mode	el 11 ***					
Delay time O							
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED				
• Non-seasonal;	no differe	encing					
Moving-average	lag ref. O 11	estimate 1.144	s.e. 0.326				
*** Transfer-fu	nction mode	el 12 ***					
Delay time O							
Transformation Constant	геf. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed				
* Non-seasonal;	no differe	encing					
Moving-average	lag ref. 0 12	estimate 1.447	s.e. 0.303				
*** Iransfer-fu	nction mode	el 13 ***					
Delay time O							
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed				
• Non-seasonal;	no differe	encing					
Moving-average	lag ref. O 13	estimate 0.00000260	s.e. 0.00000115				
*** Transfer-fu	nction mode	el 14 ***					
Delay time O							
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED				

<ul> <li>Non-seasonal;</li> </ul>	no differ	rencing	
Moving-average	lag ref. 0 14	estimate 0.00000377	s.e. 0.00000126
*** Transfer-fu	nction mod	del 15 ***	
Delay time O			
Transformation Constant	ref. D O	estimate 1.00000 0.	s.e. Fixed Fixed
* Non-seasonal;	no differ	rencing	
Movi <b>ng</b> -average	lag ref. O 15	estimate -0.00000037	s.e. 0.00000119
*** Autoregress	ive moving	-average model	***
Innovation varia	ance 0.376	53	
Transformation Constant	ref. 0 16	estimate 1.00000 -0.37	s.e. FIXED 3.48
* Non-seasonal;	no differ	rencing	
Moving-average	lag ref. 1 17	estimate -0.236	s.e. 0.104

. . . .

ł

ł

÷.

58 FORECAST [MAXLEAD=24] INTREND, INM2, INM3, INM4, INM5, \
59 INM6, INM7, INM8, INM9, INM10, INM11, INM12, FF, LAGFF, LLAGFF

59.....

\*\*\* Forecasts \*\*\*

Maximum lead time: 24

	<i>.</i>		
Lead time	forecast	lower limit	upper limit
1	4.924	3,915	5.933
2	5.709	4.673	6.746
3	6.274	5.238	7.311
4	5.777	4.741	6.814
5	5.919	4.882	6.955
6	6.071	5.035	7.108
7	5.818	4.781	6.854
8	5,990	4.953	7.026
ŷ	6.709	5 672	7 746
10	5.935	4.898	6 972
11	6 169	5 132	7 205
12	6 508	5 471	7.544
13	5 072	6 035	£ 10g
14	5 550	4.000	4 597
15	6 1/6	5 110	7 107
14	0,140 5 450	J. 110	/ . 185
10	2.000	4.015	0.000
17	5.792	4.756	6.829
18	5.946	4.909	6.982
19	5.693	4.656	6.730
20	5.866	4,829	6.903
21	6.586	5.550	7.623
22	5.813	4.777	6.850
23	6.048	5.011	7.085
24	6.388	5.351	7.424

- 60

- 61 TKEEP RESID=RES1 62 CALC N1=NOBS(RES1) 63 CALC ESS1=VAR(RES1)\*(N1-1) 64 PRINT ESS1

ESS1 33.49

- 65 CALC R21=(TSS-ESS1)/TSS
- 66 CALC R2S1=(DSSS-ESS1)/DSSS 67 PRINT R21,R2S1

R2S1 0.4698 R21 0.6158

```
68 <sup>II</sup>
```

- -69 Using an IMA(0,1,1) model for errors
  -70 "
  71 TSM ERM2; ORDERS=!(0,1,1)
  72 TRANSFER TREND,M2,M3,M4,M5,M6,M7,M8,M9,M10,M11,M12,FUEL,LAGFUEL,LLFUEL
  73 ESTIMATE[PRINT=ESTIMATES; CONSTANT=F] STDACC; TSM=ERM2

73				 	
**** Time-ser	ies analysis	****			
*** Transfer-f	unction model	1 ***			
Delay time O					
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed		
• Non-seasonal	; no differer	ncing			
Moving-average	lag ref. 0 1	estimate -0.0282	s.e. 0.0148		
*** Transfer-fi	unction model	2 ***			
Delay time O					
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED		
• Non-seasonal	; no differer	ncing			
Moving-average	lag ref. 02	estimate 0.646	s.e. 0.309		
*** Transfer-f	unction model	3 ***			
Delay time O					
Transformation Constant	ref. D O	estimate 1.00000 0.	s.e. Fixed Fixed		
* Non-seasonal	; no differen	ncing			
Moving-average	lag ref. 0 3	estimate 1.153	s.e. 0.338		
*** Transfer-f	unction mode	4 ***			
Delay time O					
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED		
* Non-seasonal	; no differe	ncing			
Moving-average	lag ref. 0 4	estimate 0.346	s.e. 0.311		

-----

ı.

\*\*\* Transfer-function model 5 \*\*\*

I

- ----

#### Delay time O

	ref.	estimate	s.e.
Transformation Constant	0	1.00000	FIXED
	Ŭ	0.	FINED
Non-seasonal;	no differe	ncing	
	lag ref.	estimate	s.e.
Moving-average	05	0.742	0.278
*** Transfer-fu	unction mode	l 6 ***	
Delay time O			
Transformation	гет. О	estimate 1.00000	S.e. FIXED
Constant	Ď	0.	FIXED
* Non-seasonal -	nn differe	cipa	
non Beasonar,			
Novina-overse	lag ref.	estimate	s.e.
Moving-average	υo	0.660	0.273
		<b>_</b>	
ransfér-tu	inction model	. / ***	
Delay time O			
	rof	antimata	
Transformation	0	1.00000	s.e. FIXED
Constant	0	0.	FIXED
• Non-seasonal:	no differer	scina	
		.eig	
Moving-oversee	lag ref.	estimate	S.e.
NOVING BACK BYC	U 7	0.639	0.275
*** 7		0 +++	
inanster-tu	nction model	0	
<b>Delay time O</b>			
	ref.	estimate	S P
Transformation	0	1.00000	FIXED
Constant	Û	0.	FIXED
* Non-seasonal;	no differen	scing	
		-	
Moving-average	lag ref. n s	estimate 0 612	s.e.
		0.012	V.ETT
*** Transfort	nation model	0 ***	
Transfer - Ju	netion model	<b>y</b>	
Delay time O			
	ref.	estimate	S.E.
Transformation	0	1.00000	FIXED
Constant	0	0.	FIXED
* Non-seasonal;	no differen	cing	
		actimete	
Moving-average	0 9	escimate 1.414	s.e. 0.276

.

\*\*\* Transfer-function model 10 \*\*\*

Delay time O

÷.

Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. Fixed Fixed
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. O 10	estimate 1.099	s.e. 0.288
*** Transfer-fu	nction mode	l 11 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. Fixed Fixed
* Non-seasonal;	no differe	encing	
Moving-average	lag ref. 0 11	estimate 1.032	s.e. 0.274
*** Transfer-fu	nction mode	el 12 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	encing	
Moving-average	lag ref. 0 12	estimate 1.483	s.e. 0.277
*** Transfer-fu	nction mode	el 13 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	encing	
Moving-average	lag ref. O 13	estimate 0.00000172	s.e. 0.00000112
*** Transfer-fu	nction mode	el 14 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED

:

ł

ī.

* Non-seasonal;	no differ	encing	
Moving-average	lag ref. 0 14	estimate 0.00000277	s.e. 0.00000118
*** Transfer-fu	nction mode	el 15 ***	
Delay time D			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
Non-seasonal;	no differe	encing	
Moving-average	lag ref. O 15	estimate -0.00000131	s.e. 0.00000115
*** Autoregress	ive moving	average model	***
Innovation varia	ance 0.3289	9	
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed
<ul> <li>Non-seasonal;</li> </ul>	differenci	ing order 1	
Moving-average	lag ref. 1 16	estimate 0.7422	s.e. 0.0758
74 EODECAST	MAVIEAD-2/	THITPEND THM	2 11/1473 13/1477 13/14

.

.

74 FORECAST [MAXLEAD=24] INTREND, INM2, INM3, INM4, INM5, \ 75 INM6, INM7, INM8, INM9, INM10, INM11, INM12, FF, LAGFF, LLAGFF

.

75.....

ł

÷

#### \*\*\* Forecasts \*\*\*

-- --

Maximum lead time: 24

Lead	time	forecast	lower limit	upper limit
	1	4.512	3.569	5.456
	2	5.209	4.234	6.183
	3	5.595	4.591	6.599
	4	5.098	4.065	6.132
	5	5.221	4.159	6.282
	6	5.385	4.296	6.474
	7	5.108	3.993	6.224
	8	5.295	4.153	6.437
	9	6.034	4.867	7.202
	10	5.285	4.092	6.477
	11	5.519	4.301	6.736
	12	5.935	4.694	7.176
	13	4.445	3.180	5,709
	14	4.868	3,581	6.156
	15	5.361	4.050	6.671
	16	4.865	3.532	6.198
	17	4.988	3.633	6.343
	18	5.152	3.775	6.529
	19	4.876	3.478	6.274
	20	5.063	3.645	6.482
	21	5.803	4.363	7.243
	22	5.054	3.594	6.514
	23	5.288	3.808	6.769
	24	5.705	4.205	7.205

- 76 TKEEP RESID=RES2
- 77 CALC N2=NOBS(RES2) 78 CALC ESS2=VAR(RES2)\*(N2-1)
- 79 PRINT ESS2

ESS2 29.27

- 80 CALC R22=(TSS-ESS2)/TSS
- 81 CALC R2S2=(DSSS-ESS2)/DSSS
- 82 PRINT R22, R2S2

R2S2 R22 0.5365 0.6642

83 "

- -84 Using an IMA(0,2,2) model for errors
- -85 "

- 86 TSM ERM3; ORDERS=!(0,2,2) 87 TRANSFER M2,M3,M4,M5,M6,M7,M8,M9,M10,M11,M12,FUEL,LAGFUEL,LLFUEL 88 ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM3

88				 	
***** Time-seri	es analysis	****			
*** Transfer-fu	unction mode	1 ***			
Delay time O					
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED		
* Non-seasonal;	no differer	ncing			
Moving-average	lag ref. 0 1	estimate 0.422	s.e. 0.311		
*** Transfer-fu	nction mode	. 2 ***			
Delay time O					
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED		
* Non-seasonal;	no differer	ncing			
Moving-average	lag ref. 0 2	estimate 0.730	s.e. 0.345		
*** Transfer-fu	nction mode	3 ***			
Delay time O					
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED		
Non-seasonal;	no differer	າວເກີ່ງ			
Moving-average	lag ref. 0 3	estimate 0.075	s.e. 0.320		
*** Transfer-fu	nction mode	L 4 ***			
Delay time O					
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED		
* Non-seasonal;	no differe	ncing			
Moving-average	lag ref. O 4	estimate 0.567	s.e. 0.284		

.

.

\*\*\* Transfer-function model 5 \*\*\*

Delay time D

Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. O 5	estimate 0.508	s.e. 0.279
*** Transfer-fu	nction mode	l 6 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 6	estimate 0.482	s.e. 0.281
*** Transfer-fu	nction mode	l 7 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 07	estimate 0.497	s.e. 0.284
*** Transfer-fu	nction mode	8 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. O 8	estimate 1.285	s.e. 0.282
*** Transfer-fu	nction model	( 9 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	ncing	
Moving-average	lag ref. 0 9	estimate 0.894	s.e. 0,293

1

i.

\*\*\* Transfer-function model 10 \*\*\*

Delay time 0

Iransformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed
• Non-seasonal;	no differ	encing	
Moving-average	lag ref. O 10	estimate 0.841	s.e. 0.279
*** Transfer-fu	nction mode	el 11 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	S.e. FIXED FIXED
• Non-seasonal;	no differe	encing	
Moving-average	lag ref. O 11	estimate 1.400	s.e. 0.278
*** Transfer-fu	nction mode	el 12 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
* Non-seasonal;	no differe	encing	
Moving-average	lag ref. 0 12	estimate 0.00000052	s.e. 0.00000115
*** Transfer-fu	nction mode	el 13 ***	
Delay time O			
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED
• Non-seasonal;	no differe	encing	
Moving-average	lag ref. O 13	estimate 0.00000034	s.e. 0.00000122
*** Transfer-fur	nction mode	l 14 ***	
Delay time O			
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. FIXED FIXED

\* Non-seasonal; no differencing

	lag	ref.	estimate	s.e.
Moving-average	Ō	14	-0.00000255	0.00000120

\*\*\* Autoregressive moving-average model \*\*\*

Innovation variance 0.3496

	ref.	estimate	s.e.
Transformation	0	1.00000	FIXED
Constant	0	0.	FIXED

• Non-seasonal; differencing order 2

	lag	ref.	estimate	s.e.
Moving-average	1	15	1.6900	0.0851
	2	16	-0.6909	0.0850

89 FORECAST [MAXLEAD=24] INM2,INM3,INM4,INM5,\
90 INM6,INM7,INM8,INM9,INM10,INM11,INM12,FF,LAGFF,LLAGFF

1

90.....

\*\*\* Forecasts \*\*\*

Maximum lead time: 24

Lead time	forecast	lower limit	upper limit
1	4.683	3.710	5.655
2	5.115	4.097	6.134
3	5.262	4.200	6.324
4	4.802	3.698	5.907
5	4.951	3.805	6.097
6	5.085	3.899	6.270
7	4.798	3.573	6.022
8	4.961	3,700	6.223
9	5.648	4.349	6.946
10	5.008	3.673	6.343
11	5.307	3,937	6 677
12	5.735	4.330	7 140
13	4,292	2.854	5 731
14	4.601	3 130	6 073
15	4 053	3 440	6 (58
16	4.707	2 057	4 070
17	4.4/2	2.751	6.030
10	4.042	3.074	0.210
10	4.//D	5.1/0	0.3/3
19	4.488	2.858	6,118
20	4.65	2.99	6.31
21	5.34	3.65	7.03
22	4.70	2.98	6.42
23	5.00	3.25	6.75
24	5.42	3.65	7.20

- 91 TKEEP RESID=RES3 92 CALC N3=NOBS(RES3) 93 CALC ESS3=VAR(RES3)\*(N3-1) 94 PRINT ESS3

ESS3 30.77

- 95 CALC R23=(ISS-ESS3)/ISS
- 96 CALC R2S3=(DSSS-ESS3)/DSSS 97 PRINT R23,R2S3

R23 R2S3 0.5129 0.6471

98 H

- -99 Using an ARIMA(0,2,2,0,1,1,12) model for errors
- -100 "
- 100 TSM ERM4; ORDERS='(0,2,2,0,1,1,12)
  102 TRANSFER FUEL,LAGFUEL,LLFUEL
  103 ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM4

103..... \*\*\*\*\* Time-series analysis \*\*\*\*\* \*\*\* Transfer-function model 1 \*\*\* Delay time O ref. estimate s.e. 1.00000 Transformation 0 FIXED Ó Ο. FIXED Constant • Non-seasonal; no differencing lag ref. estimate s.e. 0.00000113 0 1 0.00000053 Moving-average \*\*\* Transfer-function model 2 \*\*\* Delay time 0 estimate ref. s.e. Transformation 0 1.00000 FIXED 0 0. FIXED Constant Non-seasonal; no differencing lag ref. estimate s.e. Moving-average 0 2 0.00000013 0.00000119 \*\*\* Transfer-function model 3 \*\*\* Delay time 0 ref. estimate s.e. 1.00000 Transformation 0 FIXED Constant 0 0. FIXED \* Non-seasonal; no differencing lag ref. estimate 0 3 -0,00000286 s.e. 0.00000117 Moving-average \*\*\* Autoregressive moving-average model \*\*\* Innovation variance 0.4355 ref. estimate s.e. 1.00000 FIXED Transformation Ω FIXED Constant ۵ 0. • Non-seasonal; differencing order 2 estimate 1.6318 s.e. 0.0887 lag ref. Moving-average 1 4 5 2 -0.6451 0.0898

\* Seasonal; period 12; differencing order 1

Moving-average		lag	ref.	estimate	s.e.
		12	6	0.600	0.113
104	FORECAST	[MAXL	.EAD=24]	FF,LAGFF,LLAGFF	

104.... 

#### \*\*\* Forecasts \*\*\*

- -

Maximum lead time: 24

Lead	time	forecast	lower limit	upper limit
	1	4.204	3.119	5.289
	2	4.469	3.313	5.626
	3	5.111	3.882	6.339
	- 4	3.977	2.675	5.278
	5	4.209	2.835	5.584
	6	4.649	3.200	6.098
	7	4.327	2.803	5.850
	8	3.997	2.398	5.597
	9	5.14	3.46	6.82
	10	4.01	2.25	5.76
	11	4.00	2.17	5.84
	12	4.97	3.06	6.88
	13	2.90	0.74	5.05
	14	3.03	0.75	5.30
	15	3.88	1.49	6.28
	16	2.73	0.21	5.25
	17	2.95	0.31	5.59
	18	3.37	0.61	6.13
	19	3.03	0.14	5.92
	20	2.68	-0.33	5.70
	21	3.81	0.68	6.95
	22	2,66	-0.60	5.92
	23	2.64	-0.75	6.03
	24	3.59	0.08	7.10

- 105 TKEEP RESID=RES4 106 CALC N4=NOBS(RES4) 107 CALC ESS4=VAR(RES4)\*(N4-1)
- 108 PRINT ESS4

# ESS4 37.37

- 109 CALC R24=(TSS-ESS4)/TSS 110 CALC R2S4=(DSSS-ESS4)/DSSS 111 PRINT R24,R2S4

R24 R2S4 0.5713 0.4083

112 "

-113 using an ARIMA(0,1,1,0,1,1,12) model for errors -114 "

- 115 TSM ERM5; ORDERS=!(0,1,1,0,1,1,12) 116 TRANSFER FUEL,LAGFUEL,LLFUEL 117 ESTIMATE [PRINT=ESTIMATES; CONSTANT=F; MAXCYCLE=50] STDACC; TSM=ERM5

i.

ł.

117	•••••			
***** ⊺íme+seri	es analysi	s *****		
*** Transfer-f∟	Inction mod	el 1 ***		
Delay time O				
Transformation Constant	ref. Q O	estimate 1.00000 0.	s.e. Fixed Fixed	
* Non-seasonal;	no differ	encing		
Moving-average	lag ref. 0 1	estimate 0.00000176	s.e. 0.00000110	
*** Transfer-fu	nction mod	el 2 ***		
Delay time O				
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. Fixed Fixed	
* Non-seasonal;	no differ	encing		
Moving-average	lag ref. O 2	estimate 0.00000249	s.e. 0.00000116	
*** Transfer-fu	nction mod	el 3 ***		
Delay time O				
Transformation Constant	ref. 0 0	estimate 1.00000 0.	s.e. Fixed Fixed	
* Non-seasonal;	no differ	encing		
Moving-average	lag ref. 0 3	estimate -0.00000156	s.e. 0.00000112	
*** Autoregress	ive moving	-average model	***	
Innovation vari	ance 0.392	7		
Transformation Constant	ref. O O	estimate 1.00000 0.	s.e. FIXED FIXED	
* Non-seasonal;	differenc	ing order 1		
Moving-average	lag ref. 1 4	estimate 0.7079	s.e. 0.0809	
* Seasonal; per	iod 12; di	fferencing or	ler 1	
	lag ref.	estimate	s.e.	

-

.

### Moving-average 12 5 0.655 0.106

118 FORECAST [MAXLEAD=24] FF,LAGFF,LLAGFF

118.....

\*\*\* Forecasts \*\*\*

Maximum lead time: 24

Lead tim	me f	orecast	lower limit	: upper limit
	1	4.252	3,221	5.282
	2	4.803	3,730	5.877
	3	5.642	4,527	6.757
	4	4.615	3.455	5.770
	5	4.891	3.697	6.085
	Ś	5.306	4.075	6.537
	7	5.028	3.760	6.295
	8	4.807	3.505	6.110
	<u>o</u>	5 926	4 589	7 263
	10	4 914	3 547	6.284
	11	4 994	3 591	6 397
	12	5 010	484	7 354
	13	3 933	2 355	5 512
	14	4 226	2 594	5 855
	15	5 10	3 51	6.87
	16	4 16	2.44	5.80
	17	4.10	2 47	· · · · · · · · · · · · · · · · · · ·
	10	4.44	2.0/	0.22
	10	4.00	5.04	
	17	4,00	2.(	0.44
	20	4.30	2.43	0.27
	21	5.48	3.55	7.43
	22	4 47	2.47	6.46
2	23	4.55	2.51	6.58
	24	5.47	3.40	) 7.54

- 119 TKEEP RESID=RES5 120 CALC N5=NOBS(RES5) 121 CALC ESS5=VAR(RES5)\*(N5-1) 122 PRINT ESS5

## ESS5 34.52

- 123 CALC R25=(TSS-ESS5)/TSS 124 CALC R2S5=(DSSS-ESS5)/DSSS 125 PRINT R25,R2S5

R25 R2S5 0.6040 0.4534

## 126 127 STOP

\*\*\*\*\*\*\*\* End of job. Maximum of 36228 data units used at line 88 (13486 left)

# APPENDIX D

# Figures for Fatalaties per crash Analyses
1.3 1.25 Number of fatalities per fatal accident 1.2 AUSFRATE, var3 1.15 1.1 1.05 1 0 40 80 120 160 200

Time Months from April 1975 to December 1990

## Seasonal Subseries Plot



Season (Months)

FIGURE D.2

## Seasonal Subseries Plot

After differencing - lag 12



Season

FIGURE D.3



AUSFRATE.var1

Number of Fatal Traffic Accidents

FIGURE D.4