J. R. Statist. Soc. A. (1986)
149, *Part* 3, *pp.* 187-227

The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling

By A. C. HARVEY and J. DURBIN†

London School of Economics and Political Science, UK

[*Read before* the Royal Statistical Society *on Wednesday, March 15th, 1986, Professor M. J. R. Healy in the Chair*]

SUMMARY

Monthly data on road casualties in Great Britain are analysed in order to assess the effects on casualty rates of the seat belt law introduced on January 31st, 1983. Such analysis is known technically as intervention analysis. The form of intervention analysis used in this paper is based on structural time series modelling and differs in significant respects from standard intervention analysis based on *ARIMA* modelling. The relative merits of the two approaches are compared. Structural modelling intervention techniques are used to estimate the changes in casualty rates for various categories of road users following the introduction of the seat belt law.

Keywords: ROAD CASUALTIES; SEAT BELT LAW; STRUCTURAL MODELS; ARIMA MODELS; INTERVENTION ANALYSIS; KALMAN FILTER; DIAGNOSTICS; CUSUM; RECURSIVE RESIDUALS; RISK COMPENSATION

1. INTRODUCTION

The wearing of seat belts by front seat occupants of cars and light goods vehicles was made compulsory in the UK on January 31 1983. The law was introduced initially for an experimental period of three years with the intention that Parliament would consider extending the legislation before the expiry of this period. The Department of Transport undertook to monitor the effect of the law on road casualties and as part of this monitoring exercise they invited us early in 1985 to conduct an independent technical assessment of the statistical evidence.

A full description of the findings of the investigation and the methods used has been given in our report to the Department (Durbin and Harvey, 1985). The purposes of the present paper are:

(a) to provide an opportunity for public discussion of the results of our analysis of the effects of the seat belt law on road casualties: and
(b) to invite a technical debate on the methodology we used and in particular on the relative merits of structural modelling and *ARIMA* modelling for this type of investigation.

The main data we examined consisted of numbers killed and numbers seriously injured each month for various categories of road users for the period from January 1969 to December 1984. Our task was to investigate changes in casualty rates attributable to the introduction of the seat belt law. The standard way for time series analysts to tackle problems of this type in recent years has been to use Box-Tiao (1975) intervention analysis based on Box-Jenkins (1970) *ARIMA* modelling. However, we have become increasingly dissatisfied with various aspects of standard *ARIMA* modelling. In particular, we have become disenchanted with the notion that the appropriate way to deal with trend and seasonal components is to eliminate

† *Present address*: Dept of Statistical and Mathematical Sciences, London School of Economics and Political Science, Houghton St., London WC2A 2AE, UK.

0035-9238/86/149187 $2.00

them by differencing. We are also sceptical about the emphasis on stationarity of the differenced series.

Our views on time series modelling are consistent with our general attitude to statistical modelling. We believe that 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 historical development of the structural models used in this work can be traced in papers by Muth (1960), Theil and Wage (1964), Harrison (1967), Harrison and Stevens (1976), Engle (1978), Kitagawa (1981), Harvey and Todd (1983), Durbin (1984) and Harvey (1984, 1985a) together with further references given in these papers. At the time we received the invitation from the Department of Transport to analyse the road casualty data, one of us (ACH) had been developing techniques and computer programs for time series structural modelling during the previous two or three years as part of an ESRC funded research programme in econometrics at the London School of Economics. It turned out that with relatively little modification these techniques and programs could be used for the road casualties analysis and we concluded that they were better suited to the purpose than standard *ARIMA* intervention analysis. We have therefore used them throughout, though we did in fact carry out some parallel *ARIMA* analyses for comparative purposes.

The paper is organised as follows. In Section 2 we present the models used in the paper and refer briefly to the Kalman filter techniques used for fitting them. Section 3 discusses the data used in the study. In Section 4 we explain how an appropriate model is selected for a particular series and discuss the diagnostic techniques used to assess the fit of models to the data. For this purpose we illustrate by considering monthly numbers of car drivers killed and seriously injured. Section 5 considers how the effect of an intervention is assessed in structural modelling. In this application the intervention is the introduction of the seat belt law. Section 6 summarises the results obtained. In Section 7 we discuss our conclusions.

2. STRUCTURAL TIME SERIES MODELS

2.1. *The Basic Structural Model*

The starting point for the construction of structural models is the traditional representation of a time series as a sum of trend, seasonal and irregular components

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \dots, T \tag{2.1}$$

where y_t denotes the tth observation, possibly after some transformation such as the logarithmic, and μ_t, γ_t and ε_t are the trend, seasonal and irregular components. A simple special case is that in which μ_t is the deterministic linear trend $\alpha + \beta t$ and γ_t is a strictly periodic function of period s where s is the number of months per year. However, this form is of very limited application since usually some provision is needed to permit the structure to evolve over time.

The basic idea of how this can be achieved came from Muth (1960) who considered the case where there is no seasonal and the trend has no slope but the level μ_t evolves over time in a random walk, giving the model

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_t = \mu_{t-1} + \eta_t, \tag{2.2}$$

where ε_t and η_t are independent white noise terms. It is easy to show that (2.2) is equivalent to a model in which the first differences $y_t - y_{t-1} = \Delta y_t$ follow an $MA(1)$ model. Muth then showed that the one-step ahead forecasts produced by the simplest form of exponentially weighted moving average given by Holt (1957) and Winters (1960) are minimum mean-square error (*MMSE*) forecasts for observations generated by model (2.2) and hence for the equivalent *ARIMA* (0, 1, 1) model.

Subsequently, Theil and Wage (1964) extended Muth's model to include a trend that is locally linear, but where both level and slope are determined by random walks. This gives the model

$$y_t = \mu_t + \varepsilon_t \tag{2.3a}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \beta_t = \beta_{t-1} + \zeta_t. \tag{2.3b}$$

where ζ_t is a white noise disturbance term independent of ε_t and η_t.

They showed that the *MMSE* forecasts for this model are given by the first two Holt-Winters recursions for the special case $\sigma_\eta^2 = 0$. Nerlove and Wage (1964) then showed that second differences of observations generated by (2.3) follow an $MA(2)$ model.

At this point in the early or mid 1960's it appears using hindsight that the theory of *MMSE* forecasting might have developed either by extension of structural models (2.2) and (2.3) or alternatively by extension of the differencing method to deal with higher-order trends and seasonals. In this connection the following quotation from p. 207 of Nerlove and Wage (1964) is of considerable interest. "We believe, therefore, that the results of this paper illustrate a general approach to the prediction of non-stationary time series, and these are, after all, the type mainly encountered in ecnomic or management problems. Thus the paper may have a somewhat wider significance than its title or primary purpose might suggest". In the event Box and Jenkins followed the *ARIMA* route.

The key to the handling of the computations needed for structural models of great generality is the Kalman filter and it is worth noting that Kalman's (1960) paper was already available when work on *ARIMA* modelling began. Nevertheless, with the computational technology available at the time the calculations required would have been extremely burdensome. Furthermore, it was a long time after the publication of Schweppe's (1965) paper before statisticians realised that the likelihood function could be constructed for quite complex models by means of the Kalman filter.

Independent work on structural modelling was being done in the 1960's by P. J. Harrison and some of this work for non-seasonal series is described in Harrison (1967). There are various ways in which seasonal components can be incorporated into structural models and two general techniques are described by Harrison and Stevens (1976).

Our preferred technique for handling seasonality is to employ the trigonometric model

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{jt} \tag{2.4a}$$

where, with s even and $\lambda_j = 2\pi j/s$,

$$\begin{bmatrix} \gamma_{jt} \\ \gamma_{jt}^* \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} \omega_{jt} \\ \omega_{jt}^* \end{bmatrix}, j = 1, \ldots, \tfrac{1}{2}s - 1, \tag{2.4b}$$

$$\gamma_{jt} = (\cos\lambda_j)\gamma_{j,t-1} + \omega_{jt}, j = \tfrac{1}{2}s, \tag{2.4c}$$

and where the ω_{jt}'s and ω_{jt}^*'s are both $NID(0, \sigma_\omega^2)$ and are independent of each other. In a more general formulation, the variances of ω_{jt} and ω_{jt}^* are permitted to vary with j. The rationale for (2.4) is that if the disturbances ω_{jt} and ω_{jt}^* are set equal to zero the seasonal pattern is constant and (2.4) then provides updating formulae from one time period to the next for each trigonometric component of the seasonal pattern. The introduction of the disturbances permits the seasonal pattern to vary over time.

Our basic structural model (BSM) is specified by the relations (2.1), (2.3b) and (2.4). Its reduced from is obtained by taking first difference Δ followed by seasonal differences Δ_s. The resulting series $\Delta\Delta_s y_t$ is found to follow an $MA(s+1)$ process where the $s+1$ moving average coefficients are complicated functions of the variances of the disturbances σ_ε^2, σ_η^2, σ_ζ^2 and σ_ω^2.

2.2. *Explanatory Variables and Intervention Effects*

The structural model may be extended by adding exogenous explanatory variables to the right hand side of (2.1) giving

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_j x_{jt} + \varepsilon_t, \tag{2.5}$$

where x_{jt} is the value of the jth explanatory variable at time t and δ_j is its coefficient. There is no necessity for the x_{jt} series to have any stationarity properties either before or after differencing. If $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2 = 0$, the model collapses to a standard regression model with a linear time trend and seasonal dummies in addition to the explanatory variables x_{jt}.

We now consider how to use structural modelling to estimate the effect of an intervention, such as the introduction of the seat belt law, at a particular point of time $t = \tau$. Within the *ARIMA* system the appropriate technique is called intervention analysis and was introduced by Box and Tiao (1975). The ideas from which this approach was developed can easily be incorporated into the structural framework. In the simplest case we assume that the effect of the intervention occurs instantaneously at time τ and leads to an immediate change in the level of the series which remains constant at an amount λ say. Other possibilities are that there is an instantaneous response at time τ which increases or decreases subsequently or that the response builds up gradually following the intervention.

These possibilities can be accommodated by extending the model (2.5) to the form

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_j x_{jt} + \lambda w_t + \varepsilon_t \tag{2.6}$$

where we refer to w_t as the intervention variable. In the simplest case where the response in instantaneous and constant, w_t is the dummy variable defined by

$$w_t = \begin{cases} 0, & t < \tau, \\ 1, & t \geqslant \tau. \end{cases} \tag{2.7}$$

The formulation (2.6) is the model used for most of the work in this paper. There are many ways in which the model could be modified or extended. For example, claendar effects and higher order local polynomials can be allowed for, and cycles can be incorporated into the model as in Harvey (1985a). Extensions can be made to handle weekly or even daily data.

2.3. *Statistical Treatment*

Structural time series models of great apparent complexity can be efficiently fitted by powerful but straightforward general techniques. The key to these techniques is that the models can be put into state space form and then handled routinely by the Kalman filter. In its general univariate form the state space model consists of the following two equations:

$$y_t = z_t'\alpha_t + \varepsilon_t, \quad t = 1, \dots, T \tag{2.8a}$$

called the measurement or observation equation, and

$$\alpha_t = T_t \alpha_{t-1} + \eta_t, \quad t = 1, \dots, T \tag{2.8b}$$

called the transition or state equation. Here, y_t is the observation of interest, z_t is a non-stochastic vector, α_t is an $m \times 1$ state vector, T_t is a non-stochastic $m \times m$ matrix and ε_t, η_t are independent $NID(0, \sigma^2 h_t)$ and $NID(0, \sigma^2 Q_t)$ disturbances where h_t, Q_t are a scalar and an $m \times m$ matrix respectively and σ^2 is a positive scalar.

The advantage of this formulation is that there is a routine mechanical procedure for updating estimates, generally referred to as the Kalman filter. Maximum likelihood estimation can be performed either in the time domain, using the Kalman filter, or in the frequency domain, see Harvey and Peters (1984) for details and Nerlove *et al.* (1979) pp. 132–139 for a general discussion of estimation in the frequency domain.

3. DATA

The principal data used for the analyses in this paper are monthly observations of (i) numbers killed and seriously injured and (ii) numbers killed for the following categories of road users: (a) car drivers, (b) car front seat passengers, (c) car rear seat passengers, (d) pedestrians and (e) cyclists. The period covered was January 1969 to December 1984. The data refer to the whole of Great Britain. We also had data for various periods on seat belt wearing rates, indices of traffic densities for various categories of vehicles and on the real cost of petrol.

To give the reader an impression of the magnitudes involved we give in Table 1 the annual totals for 1982 and 1984 as representing two complete years before and after the completion of the seat belt law.

TABLE 1
Numbers killed and seriously injured and numbers killed 1982 *and* 1984

	Killed and seriously injured		*Killed*	
	1982	1984	1982	1984
Car drivers	19,460	16,421	1,472	1,228
Car front seat passengers	9,458	7,047	658	539
Car rear seat passengers	4,706	5,062	297	372
Pedestrians	18,963	19,168	1,869	1,821
Cyclists	5,967	6,506	294	337

4. MODEL SELECTION

In the first two sections we argued that the attraction of structural time series models is that they are formulated in terms of components which have a direct interpretation. In this section we set out what we feel is an appropriate model selection methodology for structural models both with and without explanatory variables. Some of the diagnostic checking techniques are appropriate for *ARIMA* as well as structural models. However, as we shall show at various points in the discussion, there are important differences in emphasis in structural and *ARIMA* model selection.

The introduction of intervention effects raises further issues of model selection. These issues are considered separately in Section 5. The present section concentrates on selecting an appropriate model for various casualty series based on data from January 1969 to December 1982, i.e. before the intervention took place. Our basic approach was first to fit the model for the period up to December 1981 and then to use the 1982 data for post-sample predictive tests.

4.1. *Univariate Time Series Models*

Fig. 1 shows the numbers of car drivers killed and seriously injured (*KSI*) in each month in Great Britain over the period January 1969 to December 1984. There is a clear seasonal pattern and although it could be argued that the inclusion of a slope component in the trend — β_t in (2.3) — is unnecessary, our preference is to at least start off with the more general

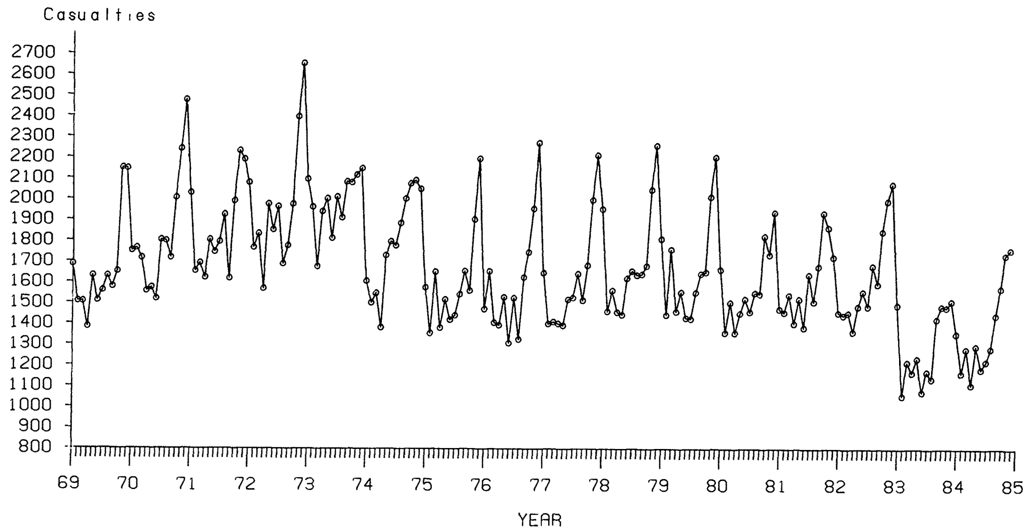

Fig. 1. Car drivers killed and seriously injured.

local linear trend model. Little is lost by proceeding in this way, but more could be lost if the slope were incorrectly constrained to be zero throughout the period.

The first model fitted was the basic structural model (*BSM*), (2.1), (2.3b) and (2.4). Estimating this model by exact *ML* for the (natural) logarithm of drivers *KSI* using data from January 1969 to December 1981 gave the following estimates:

$$\underset{(0.582\times 10^{-3})}{\tilde{\sigma}_\varepsilon^2 = 3.871\times 10^{-3}},\ \underset{(0.252\times 10^{-3})}{\tilde{\sigma}_\eta^2 = 0.609\times 10^{-3}},\ \tilde{\sigma}_\zeta^2 = 0,\ \tilde{\sigma}_\omega^2 = 0$$

with

$$\tilde{\sigma} = 0.076 \qquad R^2 = 0.76 \qquad R_s^2 = 0.27$$

$$H(47) = 1.025 \qquad Q(15) = 16.80 \qquad \text{Normality} = 1.87,$$

where the figures in parentheses under the parameter estimates are asymptotic standard errors, calculated by the frequency domain method given in Harvey and Peters (1984).

The standardized residuals from fitting the model will be denoted by $\tilde{v}_t = v_t/f_t^{1/2}$, $t = s+2, \ldots, T$, where v_t is the one step ahead prediction error and f_t is the estimate of its variance. Both v_t and f_t are obtained from the Kalman filter. The various statistics presented under the above parameter estimates are as follows:

(a) $\tilde{\sigma}^2$ is the estimated one step ahead prediction error variance,

(b) R^2 is, as usual, one minus the ratio of $(T-s-1)\,\tilde{\sigma}^2$ to the sum of squares of y_t about its mean,

(c) R_s^2 is one minus $(T-s-1)\tilde{\sigma}^2$ divided by the sum of squares of first differences around the seasonal means of first differences; see Harvey (1984, Appendix 1).

(d) H is a heteroscedasticity test statistic defined by

$$H(m) = \sum_{t=T-m+1}^{T} \tilde{v}_t^2 \Big/ \sum_{t=s+2}^{s+1+m} \tilde{v}_t^2. \tag{4.1}$$

The integer m is approximately $(T-s-1)/3$. If the relative variances $\sigma_\eta^2/\sigma_\varepsilon^2$, $\sigma_\zeta^2/\sigma_\varepsilon^2$ and $\sigma_\omega^2/\sigma_\varepsilon^2$

were known, the distribution of H would be $F_{m,m}$ under the null hypothesis that the model is correctly specified.

(e) $Q(P)$ is the Box-Ljung statistic constructed from the first P autocorrelations of the standardised residuals. Under the null hypothesis $Q(P)$ should be treated as having a χ^2 distribution with P-3 degrees of freedom. However, when σ_ω^2 and σ_ζ^2 are both zero the appropriate number of degrees of freedom is P-1.

(f) 'Normality' is the test statistic for testing normality given by

$$\text{Normality} = \frac{(T-s-1)}{6} b_1 + \frac{(T-s-1)}{24} (b_2 - 3)^2 \tag{4.2}$$

where $\sqrt{b_1}$ is the third moment of the standardised residuals about the mean and b_2 is the fourth moment; see Jarque and Bera (1980) and Bowman and Shenton (1975). The distribution of the test statistic under the null is asymptotically χ_2^2.

None of the diagnostics indicates that the model is inappropriate and the plot of the residuals, $\tilde{v}_t$, $t = s+2, \ldots, T$, showed nothing unusual.

The model was used to predict casualties in 1982, using the estimates calculated from 1969 to 1981 data. The goodness of fit in this post-sample period was then tested using the post-sample predictive test statistic

$$\xi(l) = \sum_{t=T+1}^{T+l} \tilde{v}_t^2 / l \tag{4.3}$$

which is similar to the Chow test for regression; see Harvey and Todd (1983) and Box and Tiao (1976). The statistic $\xi(l)$ approximately follows an F-distribution with $(l, T-s-1)$ degrees of freedom. For the data in question $\xi(12)$ was 0.450, which is clearly not significant. In fact it indicates rather better predictions than in the sample period.

The above analysis suggests that the basic structural model is indeed appropriate for the drivers *KSI* series. Furthermore, re-estimating the model using data up to December 1982 changes the parameter estimates very little.

The level of the trend and its slope, together with the seasonal pattern normally change over time and they could be estimated by a smoothing algorithm. The final estimates at the end of the sample are, however, produced by the Kalman filter and these are usually the figures of most interest. In fact in the present example the slope of the trend and the seasonal pattern remain constant over time. At the end of 1981 the components of the trend were estimated as follows:

$$\text{Level } (\tilde{\mu}_T): \underset{(0.036)}{7.337} \qquad \text{Slope } (\tilde{\beta}_T): \underset{(0.0020)}{-0.0005}.$$

The figures in parentheses are root mean square errors (rmse's). Bearing in mind that the observations are in logarithms, the estimated (level of the) trend at the end of 1981 was $\exp(7.337) = 1536$. The growth rate was -0.05% which is clearly insignificant.

The seasonal effects are also produced. Exponentiating these figures gives a set of multiplicative seasonal factors. For car drivers *KSI* these are

J	*F*	*M*	*A*	*M*	*J*	*J*	*A*	*S*	*O*	*N*	*D*
1.020	0.908	0.934	0.866	0.946	0.916	0.967	0.969	0.993	1.073	1.204	1.281

As can be seen the main adverse seasonal effects occur in November and December.

4.2. *Correlograms and ARIMA Model Selection*

As we have just seen, the diagnostics confirm the choice of the basic structural model as a reasonable one for the car drivers *KSI* series. In the Box-Jenkins approach, on the other hand, no prior ideas about the nature of the series are involved and the idea is to use statistics such as the correlogram to select a model in the (seasonal) *ARIMA* class. Our contention is firstly that this class is itself somewhat arbitrary, and secondly that by following the model selection approach advocated in Box and Jenkins (1970) one can be led to the selection of models within that class which have unacceptable properties.

Fig. 2 shows the correlogram of $\Delta\Delta_{12}y_t$, where y_t is the logarithm of car drivers *KSI* for January 1969 to December 1982. The first point to note is that although this correlogram was not used in the selection of a structural model, it is quite consistent with the theoretical autocorrelation function implied by the basic structural model fitted in the previous sub-section. We would like to stress that the *BSM* is only one model within the structural class, and we would not advocate fitting it if its implied autocorrelation function was incompatible with the observed correlogram.

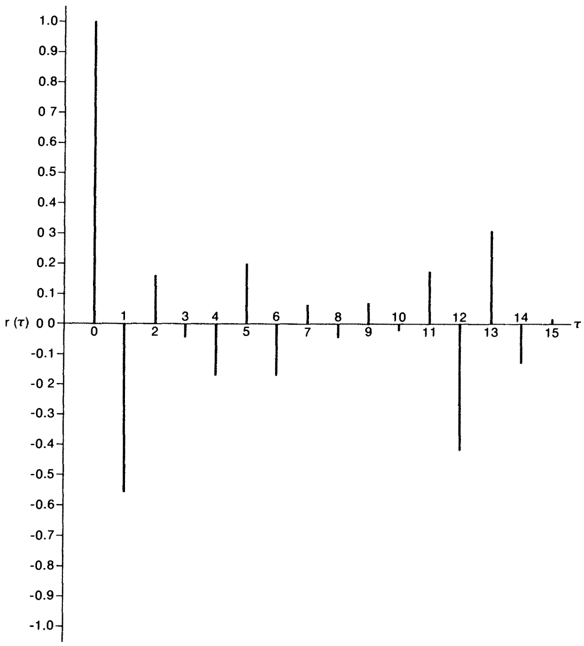

Fig. 2. Correlogram of $\Delta\Delta_{12}y_t$ for car drivers *KSI*.

As regards *ARIMA* model selection, the correlogram in Fig. 2 clearly leads to the choice of the airline model

$$\Delta\Delta_{12}y_t = (1 + \theta L)(1 + \Theta L^{12})\xi_t \tag{4.4}$$

where $L$is the lag operator. Using the data up to December 1982 to estimate such a model with the SAS package gave $\tilde{\theta} = -0.684$ and $\tilde{\Theta} = -0.995$ with $\tilde{\sigma} = 0.075$. In fact this is compatible with the *BSM* we fitted in sub-section 4.1 since the reduced form of a *BSM* with $\sigma_\zeta^2 = \sigma_\omega^2 = 0$ is an airline model with $\Theta = -1$. The parameter estimates reported for the *BSM* fitted in sub-section 4.1 (to the data up to December 1981) imply $\theta = -0.674$ in the airline model.

The prevalence of the airline model in applied work suggests that this model is appropriate for many economic time series. Indeed although the airline model is not normally equivalent to the *BSM*, as it is in the example cited here, its properties are usually fairly similar; see Maravall (1985). However, despite the appeal of the airline model it is not clear that an *ARIMA* model builder would necessarily select it in this case. The reason is that when Θ is minus one, the model is strictly noninvertible. The $\Delta\Delta_{12}$ transformation then leads to overdifferencing since only the Δ_{12} operator is needed to make the series stationary.

Fig. 3 shows the correlogram of $\Delta_{12}y_t$. (The broken lines should be ignored for the moment). Although the sample autocorrelations die away in a manner compatible with a stationary series, it is not clear what *ARIMA* model would be selected. There is a fairly wide range of possibilities involving various mixtures of *AR*, *MA* and seasonal *MA* polynomials. However, it is unlikely that any of these would be a suitable parsimonious model. If the data really are generated by an airline model with $\Theta = -1$, or equivalently by a *BSM* with $\sigma_\zeta^2 = \sigma_\omega^2 = 0$, then

$$\Delta_{12}y_t = (1 + \theta L)S(L)\xi_t \tag{4.5a}$$

$$= S(L)\eta_t + \Delta_{12}\varepsilon_t, \tag{4.5b}$$

where $S(L) = 1 + L + \ldots + L^{s-1}$. Setting $\theta = -0.684$ in (4.5a) leads to the implied autocorrelation function shown by the broken lines in Fig. 3.

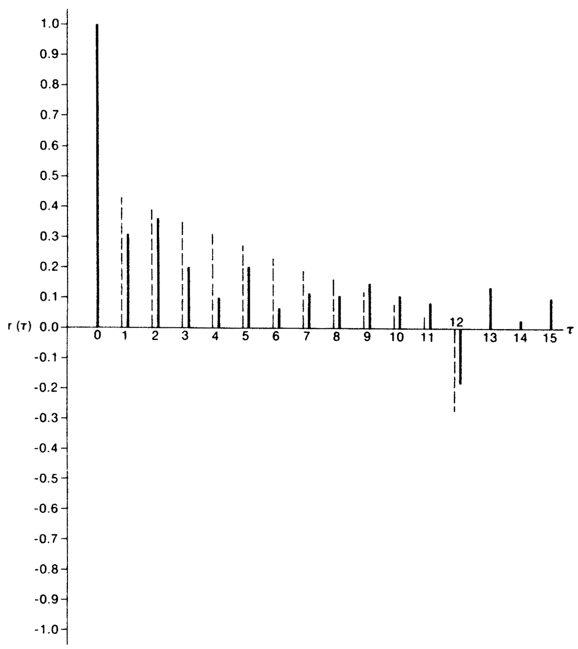

Fig. 3. Correlogram of $\Delta_{12}y_t$ for car drivers KSI.

The use of the seasonal difference operator by itself is not uncommon in *ARIMA* modelling. However an autocorrelation function such as that displayed in Fig. 3 suggests that seasonal differencing only would potentially lead to a wide range of inappropriate *ARIMA* models being chosen. As a specific case we cite the study by Bhattacharyya and Layton (1979). Their work is an investigation, using quarterly data, of the effects of the seat belt law in Queensland on deaths of car drivers. It represents a well carried out example of the use of *ARIMA* and intervention analysis techniques. The authors found that the Δ_4 operator was sufficient to reduce the observations to stationarity and on the basis of the correlogram of $\Delta_4 y_t$ they selected the following model for the pre-intervention period:

$$\Delta_4 y_t = \theta_0 + (1 + \theta_4 L^4)(1 + \theta_3 L^3 + \theta_5 L^5)\xi_t. \tag{4.6}$$

The properties of this model are not particularly appealing and the fact that Bhattacharyya and Layton report that it failed the Chow test (albeit marginally) could well be an indication of its overparameterization and inappropriateness.

A transformation which we have found to be useful for analysing economic time series is to take first differences and then to remove the seasonal means. The value of this transformation was implicitly recognised by Pierce (1978) and, as already noted in the previous sub-section, it is the sum of squares from this transformation which provides the baseline for the R_s^2 measure of goodness of fit. Applying this transformation to the car drivers *KSI* series gives the correlogram shown in Fig. 4. The dotted horizontal lines show ± 2 s.d.'s on the assumption that the first-order autocorrelation is the only non-zero one and its theoretical value is equal to the sample value of $r(1) = -0.475$. In contrast to the correlogram of $\Delta_s y_t$, the message in Fig. 4 is clear. It indicates an $MA(1)$ process and this is consistent with the fitted basic structural and airline models.

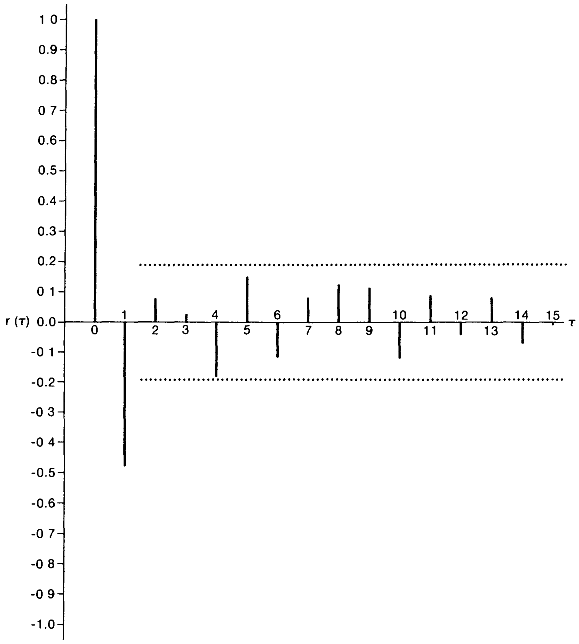

Fig. 4. Correlogram of Δy_t minus seasonal means for car drivers KSI.

4.3. *Explanatory Variables*

Suppose that it is felt that some of the movements in the series can be accounted for by observable exogenous explanatory variables. This leads to consideration of a model of the form (2.7). If $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2 = 0$, then the model collapses to a standard regression with a linear time trend and seasonal dummies in addition to the x_j's. This was the model used by Scott and Willis (1985) in their analysis of the effects of the seat belt law. However, assuming such a structure at the outset can result in misleading inferences being drawn if the assumption that σ_η^2, σ_ζ^2 and σ_ω^2 are zero is not true; see Nelson and Kang (1984).

As in most areas of the social sciences, there is no firm theory indicating which explanatory variables should appear in a model. It is therefore important to adopt a methodology which guards against what is known as 'data mining'. This term refers to uncritical examination of data sets for evidence of relationships without adequate prior consideration of potential behavioural hypotheses and without proper allowance for the process of selection of relationships when carrying out statistical tests. The methodology we use here involves formulating a general model which includes all potentially useful explanatory variables, and then testing down to obtain a parsimonious representation. At all stages the model is checked using the diagnostics described in the previous section and the coefficients are examined to see that they are consistent with what any theory would suggest.

In modelling the series on car drivers *KSI* two explanatory variables are available for inclusion. These are:

(i) the car traffic index, which measures the number of kilometres travelled by cars in a month, and

(ii) the real price of petrol, i.e. the price of petrol per litre at the pump divided by the retail price index.

The use of the car traffic index as an explanatory variable calls for no explanation, but perhaps a little should be said about the inclusion of the real price of petrol. During the period under review there were substantial changes in this variable and one effect of petrol price increases was probably to induce some drivers to drive more slowly and with less braking and acceleration. This could be expected to reduce accident rates. The variable can also be regarded as a proxy for such factors as petrol rationing and the introduction of lower speed limits during the oil crisis period of 1973–74. Lagged values of the petrol price index were originally included but their coefficients were found to be small and statistically insignificant.

Estimating (2.5) for the period January 1969 to December 1981 by exact *ML* in the time domain with the logarithms of both the above variables included gave the following results:

$$\tilde{\sigma}_\varepsilon^2 = 4.198 \times 10^{-3}, \tilde{\sigma}_\eta^2 = 0.308 \times 10^{-3}, \tilde{\sigma}_\zeta^2 = 0, \tilde{\sigma}_\omega^2 = 0.$$

$$\text{Car traffic index coefficient} = \underset{(0.14)}{0.08}, \quad \text{Petrol price coefficient} = \underset{(0.11)}{-0.31}$$

with

$$\tilde{\sigma} = 0.074 \qquad R^2 = 0.78 \qquad R_s^2 = 0.31$$

$$H(47) = 1.43 \qquad Q(15) = 17.35 \qquad \text{Normality} = 1.19.$$

Because the variables are in logarithms the coefficients of the car traffic index and the petrol price may be interpreted as elasticities. Thus a 1% rise in the traffic index gives a 0.08% rise in casualties, while a 1% rise in the price of petrol gives a 0.31% fall in casualties. The coefficient of the traffic index is statistically insignificant, and when we refitted the model without it, we obtained virtually the same results for the other parameters; see Durbin and Harvey (1985, p.A7). The diagnostics were satisfactory and the post sample predictive test statistic is $\xi(12) = 0.562$ indicating that the model gives good predictions for 1982.

On a technical point it should be noted that the diagnostics H, Q and Normality for a model of the form (2.7) are constructed from the residuals obtained by applying the Kalman filter to the 'observations' $y_t - x_t'\tilde{\delta}$, $t = 1, \ldots, T$. This Kalman filter is the same as the Kalman filter appropriate for the *BSM* and so the number of residuals is $T - s - 1$. However, if the explanatory variable coefficient vector, δ, is included in an augmented state vector only $T - s - 1 - k$ residuals are obtained. These residuals are known as generalized recursive residuals; see Harvey and Peters (1984). When explanatory variables are present in the model, the Chow statistic, (4.3), is computed using generalized recursive residuals.

4.4. *Detecting Model Breakdown*

The use of CUSUMs of recursive residuals for detecting structural change over time in linear regression models was first suggested by Brown, Durbin and Evans (1975). Similar techniques can also be used to see whether a structural time series model with explanatory variables is breaking down. The residuals used are the generalized recursive residuals described at the end of the previous sub-section.

Let $\tilde{v}_t^\dagger$, $t = s + k + 2, \ldots, T$ denoted the standardized generalized recursive residuals. Then

$$CUSUM(t, \tau) = \sum_{j=\tau+1}^{t} \tilde{v}_j^\dagger, \; t = \tau + 1, \ldots, T. \tag{4.7}$$

If the relative variances $\sigma_\eta^2/\sigma_\varepsilon^2$, $\sigma_\zeta^2/\sigma_\varepsilon^2$ and $\sigma_\omega^2/\sigma_\varepsilon^2$ were known, the generalized recursive residuals would have the same properties as the recursive residuals in linear regression, i.e. they would be normally and independently distributed with mean zero and constant variance. Fig. 5 shows an example of a *CUSUM* for the model fitted to pedestrians killed, the variance parameters having been estimated using data from January 1969 to December 1982. The steady rise in the *CUSUM* from the beginning of 1982 onwards suggests that the model may have started to break down at this point, since it was systematically underpredicting. The fact that the significance lines are not crossed is not critical since the *CUSUM* is best regarded as a diagnostic rather than a formal test procedure.

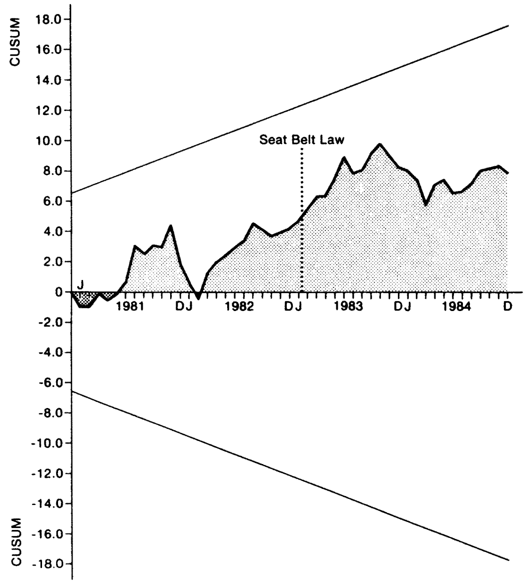

Fig. 5. Cusum of standardised recursive residuals for pedestrians killed from January 1981 onwards with 10 per cent significance lines.

A second example of the use of a *CUSUM* is given in Fig. 6. This shows the *CUSUM* for rear seat passengers killed starting at $\tau =$ January 1983. In this case the *CUSUM* line soon crosses the significance line, clearly indicating that the model began to break down in early 1983.

As already indicated, the *CUSUM* is best regarded as a diagnostic rather than a formal test, particularly when the investigator has no prior knowledge as to the point at which a model is liable to break down. However for investigation of change after a given value of τ, there are other more formal and possibly more powerful tests which may be applied. Two possibilities are (a) the recursive residual t-test which was suggested by Harvey in the discussion to Brown, Durbin and Evans (1975, p. 179) and (b) the Chow test based on (4.3).

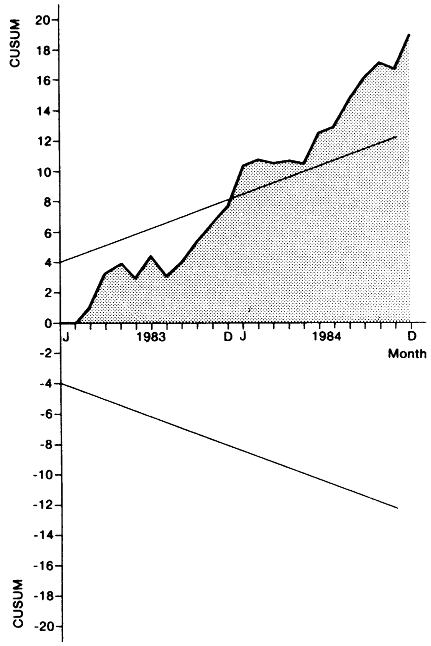

Fig. 6. Plot of cusum of standardised residuals for car rear seat passengers killed from February 1983 onwards with 10 per cent significance lines (intervention variable not included in the model).

5. ASSESSING THE EFFECTS OF INTERVENTIONS

The effect of an intervention on the series of interest can be assessed by including a variable, w_t, in the model as in (2.6). The actual form of w_t will depend on the effect which the intervention is assumed to have on the series. The final effect may involve a shift in the level or the slope of the trend, or a combination of both. Alternatively the effect may be transient. In addition there is the question of the dynamic pattern of the response.

The use of intervention variables introduces features into estimation which raise special problems for model selection. These arise because the intervention is normally a once and for all event. In subsection 5.1 we discuss estimation in some detail before suggesting a model selection strategy. The general conclusions apply irrespective of whether or not explanatory variables are present in the model.

5.1. *Estimation*

Suppose that the relative variances $\sigma_\eta^2/\sigma_\varepsilon^2$, $\sigma_\zeta^2/\sigma_\varepsilon^2$ and $\sigma_\omega^2/\sigma_\varepsilon^2$ are known or, alternatively, that the sample size prior to the intervention is large so that they can effectively be treated as known. Unless the trend and seasonal components in (2.6) are all deterministic, i.e. $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2 = 0$, the following results hold for the estimator of λ:

(a) the influence of observations beyond the intervention tends to zero as the time beyond the intervention tends to infinity; and
(b) the estimator is not consistent, even though it is the *MVUE*.

In order to explore the implications of the above results further, consider the simple random walk plus noise model, (2.2), with a step intervention, (2.7), at time $t = \tau$, i.e.

$$y_t = \mu_t + \lambda w_t + \varepsilon_t,\ \mu_t = \mu_{t-1} + \eta_t, \quad t = 1, \ldots, T \tag{5.1}$$

with $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, $\eta_t \sim NID(0, \sigma_\eta^2)$ and $\sigma_\eta^2/\sigma_\varepsilon^2 = q \geqslant 0$. The reduced form of the random walk plus noise model is *ARIMA* (0, 1, 1), i.e.

$$\Delta y_t = \xi_t + \theta\xi_{t-1}, \quad \xi_t \sim NID(0, \sigma^2), \tag{5.2}$$

with

$$\theta = -2/\{2 + q + \sqrt{(q^2 + 4q)}\} \qquad \text{and} \qquad \sigma^2 = -\sigma_\varepsilon^2/\theta. \tag{5.3}$$

It is not difficult to see that for a given value of q the variance of the *GLS* estimator of λ is

$$\operatorname{var}(\tilde{\lambda}) = \sigma^2\{1 - (-\theta)^2\}/\{1 - (-\theta)^{2(T-\tau+1)}\}. \tag{5.4}$$

In the limit as $T - \tau \to \infty$

$$\operatorname{var}(\tilde{\lambda}) \underset{T-\tau\to\infty}{\to} \sigma^2\{1 - (-\theta)^2\} \tag{5.5}$$

since for $0 < q \leqslant \infty$, $-1 < \theta \leqslant 0$. Thus λ cannot be estimated consistently unless $q = 0$ in which case $\theta = -1$ and $\operatorname{var}(\tilde{\lambda}) = \sigma^2/(T - \tau + 1)$.

It is interesting to compare the variance of the estimator of λ based on a finite number of observations after the intervention with its variance for an infinite number of observations after the intervention. Thus for $q = 0.1$ we have the following relative variances for different values of $l = T - \tau + 1$:

l	1	2	3	4	5	6	7	12
Rel. var	.47	.72	.85	.92	.96	.98	.99 ...	1

A more extensive table can be found in Harvey (1985b).

The above comparisons are basically applicable to the models for car drivers *KSI* reported in Section 4 for large sample sizes. For the basic structural model $q = 0.157$, but when explanatory variables were added q fell to $q = 0.080$. In both cases q is close enough to 0.1 to make the table above informative. As can be seen it is the first four or five observations after the intervention which contain most of the relevant information.

During our study the question of whether to use annual or monthly data was raised with us a number of times. There seemed to be a feeling that results from using annual totals might be more reliable because the annual observations tend to average out the irregularities in the monthly observations. In fact this is not an argument for using annual data and it is worth considering briefly the loss in efficiency which arises in large samples when estimating a step intervention effect from annual, rather than monthly, observations. Again we use model (5.1). As shown in Tiao (1972), the reduced form of the model is still *ARIMA* (0, 1, 1) although now the range of the moving average parameter is $[-1, 0.268]$. For $q = 0.1$ the loss of efficiency from using annual observations is approximately one third. On the other hand if $q = \infty$, i.e. $\sigma_\varepsilon^2 = 0$ and the monthly observations follow a random walk, the efficiency of the annual estimator is only 0.14.

The main conclusion from this sub-section is that discounting can be quite considerable. The solution is to try and find exogenous variables, possibly control variables, which reduce the variation in the nonstationary stochastic part of the model. Thus in terms of (5.1) we would ideally like to find an exogenous variable to add to the right hand side of (5.1a), the effect of which would be to send q to zero. The theory underlying the use of control variables as explanatory variables in equations like (2.6) is developed in Harvey (1985b).

5.2. *Model Selection*

The model selection problem surrounding the specification of an intervention effect has special features because the event only occurs once. This makes it virtually impossible to set up a general model for the intervention effect involving, say, slope, level and transient effects each with a lag structure, estimating such a model and then testing down to obtain a parsimonious specification. Instead it is necessary to proceed by specifying an intervention model on the basis of *a priori* considerations. Special diagnostics are then used to check the model. If *a priori* considerations suggest more than one possible specification, a choice between them can be made on their ability to satisfy the diagnostics and their goodness of fit. Clearly it is the pattern of residuals immediately after the intervention which is of prime importance. Several diagnostic tests are developed below. It is assumed that a fairly long series is available prior to the intervention and that this series has been used to select and estimate a suitable model. If this is not the case then some modification to the overall strategy is necessary, although the general principle remains the same.

The diagnostics we suggest are all based on generalized recursive residuals. Thus the explanatory variables are included in the state vector. In addition the intervention variable is also included in the state vector. However, it is only brought into the state vector at time τ. As a result, the prediction error at time τ is identically equal to zero; cf Brown, Durbin and Evans (1975, p. 152–3). This fact must be borne in mind when constructing test statistics. Our suggested diagnostics, apart from an examination of the plot of the residuals in the post intervention period, are as follows.

(a) *Post intervention predictive test*—The test statistic has a similar form to the Chow statistic, (4.3), i.e. the post-sample predictive test statistic. The sample size is taken to be τ-1 while the post-intervention period consists of the observations at time $t = \tau + 1, \tau + 2, \ldots, \tau + l$. The statistic is

$$\xi_\tau(l) = \sum_{t=\tau+1}^{\tau+l} \tilde{v}_t^{\dagger 2}/l \tag{5.6}$$

It is approximately distributed as $F(l, \tau\text{-}2\text{-}s\text{-}k)$,

(b) *Recursive t-test*—If *a priori* theory suggests that the residuals after the intervention may all be of a certain sign, a recursive t-test could be appropriate. The test statistic is defined as

$$\psi(l) = l^{-1/2} \sum_{t=\tau+1}^{\tau+l} \tilde{v}_t^{\dagger}. \tag{5.7}$$

This is distributed approximately as Student's t with τ-2-s-k-degrees of freedom.

(c) *CUSUM test*—The post-intervention *CUSUM* is

$$CUSUM(h) = \sum_{t=\tau+1}^{\tau+h} \tilde{v}_t^{\dagger}, h = 1, \ldots, l. \tag{5.8}$$

Boundary lines can be constructed as for the *CUSUM* defined in Section 4. As before a one-sided boundary may be appropriate.

5.3. *The Effect of the Seat Belt Law on Car Drivers*

We now turn to the application of these techniques to the series of car drivers *KSI* and car drivers killed. The most straightforward hypothesis to adopt is that the introduction of the seat belt law on January 31st, 1983, induced a once and for all downward shift in the level of each series. This implies a model of the form (2.6) with w_t defined by (2.7). However, since the seat belt wearing rate rose from 40% in December 1982 to 50% in January 1983 in anticipation of the introduction of the law, we modified (2.7) slightly by setting $w_t = 0.18$ in January 1983.

The model for car drivers *KSI* was developed earlier in Section 4.3. The relative variances were re-estimated using data up to the end of 1982. The estimates did not change very much and it was these estimates which were used when the intervention parameter, λ, was estimated using the data up to and including December 1984. The resulting estimate of λ was

$$\tilde{\lambda} = \underset{(0.053)}{-0.262.}$$

Since $\exp(-0.262) = 0.770$, the estimated reduction in drivers *KSI* is 23.0%. The 50% confidence interval for this reduction is 20.2 to 25.8% while the 95% confidence interval is 14.7% to 30.6%. The relationship between the normal and lognormal distributions suggests the use of $\exp\{\tilde{\lambda} + \frac{1}{2}\operatorname{var}(\tilde{\lambda})\}$ instead of $\exp(\tilde{\lambda})$. However the value of this gives an estimated reduction of 22.9% so the difference is negligible.

The diagnostics are based on the residuals for February 1983 onwards, i.e. τ = Jan 1983. The post intervention predictive test statistics for $l = 3$, 6 and 23 were as follows:

$$\xi_\tau(3) = 1.28, \quad \xi_\tau(6) = 0.93, \quad \xi_\tau(23) = 0.67.$$

None of these indicates a statistically significant increase in the variance of the residuals following the intervention. Similarly the *CUSUM* up to Dec. 1984, which can be found plotted in Figure 7, does not cross either the 5% or 10% boundary lines. Finally Figure 8 shows the predictions for 1983 and 1984. These are obtained using observations of the explanatory variables for 1983 and 1984 but not using observations of the car drivers *KSI* series itself for 1983 and 1984. Including the intervention variable which was subsequently estimated from the 1983 and 1984 data gives very accurate results.

5.4. *Sensitivity Analysis*

One of the attractions of structural time series models is that they lend themselves to sensitivity analysis. As an example of such analysis, consider the series for the logarithm of pedestrians killed by cars using the logarithm of pedestrians *KSI* by heavy goods vehicles (HGVs) and public service vehicles (PSVs) as an explanatory variable. (The reason for the choice of this variable as a control variable is discussed in more detail in Durbin and Harvey (1985); the details are not important in the present context.) The result of fitting this model to data up to Dec 1982 was to produce deterministic trend and seasonal components, i.e. $\sigma_\eta^2 = \sigma_\zeta^2 = \sigma_\omega^2 = 0$. The diagnostics were acceptable and $\tilde{\sigma} = 0.149$ while $R_s^2 = 0.44$ and $R^2 = 0.76$.

In view of our discussion in Section 2 of this paper, a deterministic trend must be regarded as being somewhat unusual and needs to be handled with some care. Thus some caution is needed in assessing the estimate of the intervention effect of the seat belt law which is

$$\tilde{\lambda} = \underset{(0.066)}{0.133}$$

and is statistically significant at the 5% level in a two-sided test. We therefore decided to examine the sensitivity of this estimate to changes in the variance parameters. The key parameter in this respect is σ_η^2, since in most of the models we fitted this tended to be significantly different from zero. Setting $\sigma_\eta^2/\sigma_\varepsilon^2 = 0.1$ yields

$$\tilde{\lambda} = \underset{(0.098)}{0.095}$$

Making this modification therefore yields an estimate smaller than the original estimate and this new estimate is not statistically significant at any reasonable level. Nevertheless it does indicate a non-negligible increase in this series and so we conclude that our original result is fairly robust.

Another aspect of sensitivity analysis concerns the treatment of outliers. There is no general agreement amongst statisticians on how to deal with outliers. In the present context observations with relatively large residuals are often associated with extreme weather conditions, as in the exceptionally cold month of December 1981. One way of dealing with the problem is to estimate models both including and excluding such outlying observations in order to determine whether the results are unduly sensitive to their presence. Handling a missing observation by the Kalman filter is straightforward. We fitted a number of our models with the December 1981 value included and excluded but the differences were negligible.

5.5 *Dynamic Response*

The effect of an intervention may be dynamic, that is the response to the intervention changes over time. Determining the form of the response in the absence of any *a priori* information is difficult because the intervention only happens once. The most satisfactory way to proceed it to postulate a pattern for the response based on theoretical grounds. The methods described in sub-section 5.2 can then be used to test whether the chosen specification is acceptable.

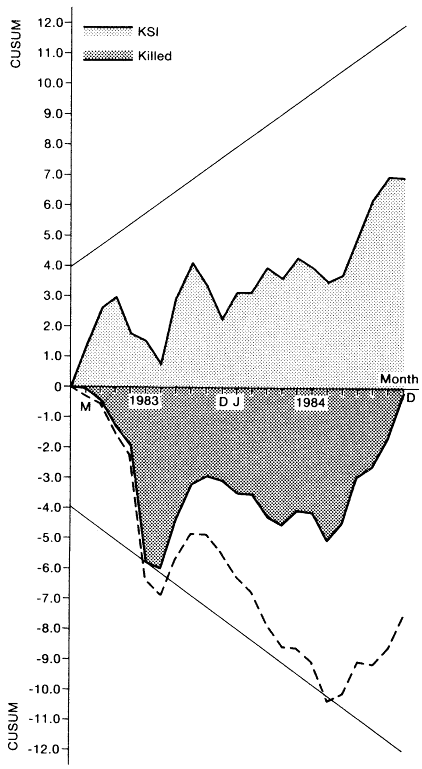

Fig. 7. Plot of *CUSUM* of standardised residuals for car drivers *KSI* and killed from March 1983 onwards with 10 per cent significance lines. Lower broken line shows CUSUM for drivers killed under the risk homeostasis hypothesis.

In the context of the seat belt law, the risk homeostasis hypothesis advocated by Wilde (1982) states that car drivers will eventually re-adjust their driving behaviour so as to keep the probability of their being killed constant. Stated in this general form the risk homeostasis hypothesis is virtually impossible to prove or disprove for a time series with a stochastic trend component. The reason is the discounting of observations as one moves further away from the time at which the intervention occured; see sub-section 5.1. However, if a specific hypothesis is put forward it can be tested. Thus suppose that the effect of the seat belt law declines linearly until it is eliminated after two years. This means that the intervention variable w_t is defined as

$$w_t = \begin{cases} 0, & t < \text{Jan. '83} \\ 0.18, & t = \text{Jan. '83} \\ 1-(t\text{-Feb. '83})/24, & t = \text{Feb. '83}, \ldots, (\text{Feb. '83}) + 24. \end{cases} \tag{5.19}$$

Fitting (2.6) to the series of car drivers killed with w_t defined by (5.9) gave

$$\tilde{\lambda} = \underset{(0.075)}{-0.206}$$

with the post-intervention test statistic $\xi_\tau(23)$, taking the value 1.240. This is not statistically significant at any conventional level of significance, and it is not much higher than the value of 1.207 obtained when the original intervention variable was fitted. Thus although one would prefer the original model on grounds of goodness of fit in the post-intervention period, the fit is not dramatically better than that of the postulated risk homeostasis model and the forecast errors from the latter model are comparable to the forecast errors obtained before the intervention. However, where this particular version of the dynamic risk homeostasis model does break down is in the *CUSUM*. Fig. 7 shows the *CUSUM* as a broken line. As can be seen it clearly crosses the lower 10% significance line. Again a one sided test is appropriate as one would expect the dynamic risk homeostasis model to overpredict if the seat belt law had led to a once and for all downward shift in the level of fatalities.

6. RESULTS

The methodology in the previous two sections was used to select models for the categories of road users we considered. This involved trials of different combinations of explanatory variables and use of the square root transformations in appropriate cases as an alternative to the logarithmic transformation, as well as an examination of residuals and inspection of the results of various diagnostic tests. In the event the logarithmic transformation was adopted for all the cases considered in this paper. The results achieved for the models we finally selected are summarised in Table 2. For each category of road user we give the estimated percentage change in numbers killed and seriously injured, and numbers killed, attributed to the introduction of the seat belt law, together with 50% confidence limits and values of diagnostic statistics.

We shall comment on the results in detail in the next section. Here, we merely draw attention to a few points of special interest. The most surprising feature of the results is the large increase of 26.7 per cent in the numbers of rear seat passengers killed. Because this increase is so much larger than expected we checked the analysis rather carefully from several points of view. For example, we put the series through the univariate *ARIMA* identification and estimation procedure and emerged with the fitted airline model

$$\Delta\Delta_{12} y_t = \underset{(0.081)}{0.212} w_t + (1\text{-}0.96L)(1\text{-}0.89L^{12})\xi_t,$$

This indicated an increase of 23.6 per cent which is close to that estimated by the corresponding structural model.

The largest departure from normality occurred for pedestrians *KSI* and this suggests the possibility of outliers. A plot of the residuals indicated two outliers with normalised values of 3.02 and -3.53 respectively. However, we concluded that these values were not inconsistent with the distribution of the full set of residuals and decided not to make any correction. A similar outcome was found for cyclists *KSI* and the same decision was taken.

Deficiencies in the fit of the model for pedestrians killed are indicated by the significantly high value of 27.2 for the Box-Ljung statistic. The fit of the model up to 1981 showed that, after

TABLE 2

Percentage changes in casualty rates and values of diagnostic statistics

	Percentage increase	*50% confidence limits*		R^2	R_s^2	*Heteroscedasticity* H	*Box-Ljung* Q	*Normality*	*Explanatory variables included*
Drivers *KSI*	−23.0	−25.8,	−20.2	0.78	0.31	1.54	16.7	1.15	*P*
Drivers killed	−18.0	−21.6,	−14.3	0.52	0.34	0.97	18.9	3.55	*TP*
FSPs *KSI*	−30.3	−33.1,	27.5	0.75	0.19	0.71	9.3	1.76	*TP*
FSPs killed	−25.1	−28.0,	−21.9	0.49	0.50	1.25	11.1	2.26	*TP*
RSPs *KSI*	2.9	−0.4,	6.4	0.81	0.50	1.10	10.5	4.39	*TP*
RSPs killed	26.7	21.5,	32.2	0.45	0.45	0.80	7.3	3.37	*TP*
Pedestrians *KSI*	−0.5	−2.3,	2.3	0.89	0.24	0.79	11.9	13.1	*T**
Pedestrians killed	7.8	4.1,	11.2	0.85	0.53	1.19	27.2	2.66	*T**
Cyclists *KSI*	4.8	0.7,	8.7	0.84	0.40	0.63	10.3	8.52	*T*C*
Cyclists killed	13.4	6.0,	21.3	0.35	0.39	0.95	17.4	2.49	*T*C*
5% significance points						0.57/1.75	21.0	5.99	

FSP = Front seat passenger
RSP = Rear seat passenger

T = Car traffic index
T^* = Total motor traffic index

P = Petrol price index
C = Cycle traffic index

allowing for the increase in road traffic, pedestrian fatalities were decreasing at a steady rate of 6.9 per cent per annum. However, the numbers for 1981 and 1982 were almost the same at 1874 and 1869. These facts, together with the examination of plots of the residuals and the *CUSUM* of the standardised recursive residuals given in Fig. 5, suggested that the decline in pedestrian fatalities which had been taking place during the 1970's may have started to level out in 1982. The effect of this would be that the model would tend to underpredict in the post-law period, thus leading to an overestimate of the effect of the law. A further point is that the standard error of the estimate of the intervention coefficient in 0.051 while the estimate of $\tilde{\lambda}$ itself is 0.075. Thus for a one-sided test $\tilde{\lambda}$ is not significant at the 5% level although it is significant at the 10% level. These reservations should be borne in mind when interpreting the estimated increase of 7.8% in pedestrian fatalities.

The values of 0.35 for R^2 and 0.39 for R_s^2 indicate that the explanatory power of the model for cyclists killed is relatively poor. On the other hand the diagnostic tests gave satisfactory results. The estimated intervention coefficient $\tilde{\lambda}$, is 0.126 with an estimated standard error of 0.100. Thus $\tilde{\lambda}$ is almost significant at the 10% level. We attribute the relative lack of precision in the estimation of λ as due more to the intrinsic variability of the series than to any deficiencies of the model.

The coefficients of the explanatory variables used in the equations reported in Table 2 provide some interesting subsidiary results. The most striking of these is that the car traffic index appears to have little effect on car drivers *KSI*, but it does affect front seat passengers and it affects rear seat passengers even more. The estimated elasticities for front and rear seat passengers are 0.33 and 0.75 respectively, with estimated standard errors of 0.16 and 0.18. For the fatality series on drivers, front seat passengers and rear seat passengers the figures are 0.32, 0.68 and 1.09 respectively. Thus the car traffic index has more of an effect on the killed series, but the ordering between the three types of car occupants remains the same. The analysis of light goods vehicles casualty series showed a similar pattern. As regards the petrol price index we found an elasticity of around minus 0.3 for the *KSI* and killed series for drivers and front seat passengers, while the elasticities for rear seat passengers *KSI* and killed were small and statistically insignificant.

The estimated elasticities of pedestrians *KSI* and killed with respect to the total motor traffic

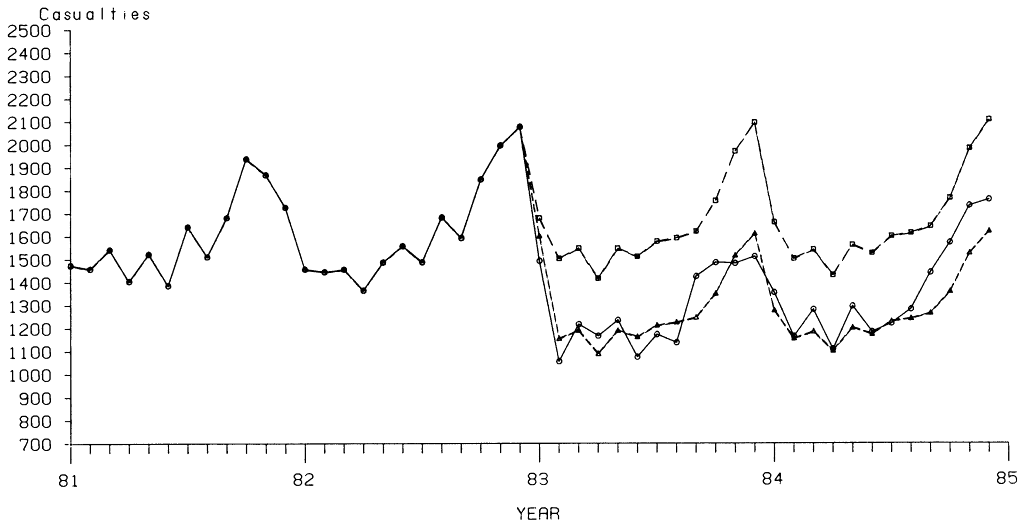

Fig. 8. Car drivers killed and seriously injured, ○, actual values of series (as in Fig. 1); △, predictions with intervention effect included; □, predictions with intervention effect removed.

index are 0.62 and 0.93 respectively. For cyclists *KSI* we found elasticities of 0.77 for the total motor traffic index and 0.14 for the cycle index. For cyclists killed the corresponding figures are 1.12 and 0.05. Thus the motor traffic index seems to be a much more important determinant of cyclist casualties than is the cycle index.

Substantially more information about our analyses and the results than space permits here is given in Appendix 1 of Durbin and Harvey (1985).

7. CONCLUSIONS

7.1. *Discussion of Results*

We first note the high rate of compliance with the seat belt law. By February 1983 the wearing rate had jumped to 90 per cent and the rate has remained at approximately 95 per cent from March 1983 onwards. There can therefore be no doubt about the success of the law as regards compliance.

In considering the casualty figures we distinguish between those directly affected by the law, namely car drivers and front seat passengers, and those not directly affected by the law, that is car rear seat passengers, pedestrians and cyclists.

Taking first numbers killed and seriously injured, we found a reduction of 23 per cent for car drivers and 30 per cent for front seat passengers. Thus for those directly affected by the law, there have been substantial reductions. For rear seat passengers *KSI* we found a rise of 3 per cent, for pedestrians a fall of one half per cent, and for cyclists an increase of 5 per cent, all three values being statistically insignificant. We conclude that there is no significant evidence of change in numbers *KSI* of those not directly affected.

For numbers killed, we found for those directly affected a reduction of 18 per cent for car drivers and 25 per cent for front seat passengers. While these reductions are not as large as for numbers *KSI* they remain substantial. Taking now those indirectly affected by the law, our model gave an increase of 27 per cent for rear seat passengers, 8 per cent for pedestrians and 13 per cent for cyclists. The value for rear seat passengers is highly significant and the other two values are on the borderline of significance. We conclude that there was an increase in fatalities of those not directly affected.

We are unable to provide a completely satisfactory explanation of the difference between the figures for *KSI* and killed for rear seat passengers, pedestrians and cyclists. As has been indicated, we found that the performance of our models for pedestrians and cyclists killed was not as good as with other data sets. The fact remains that we find the large proportionate increase in rear seat passengers killed hard to understand. We have ruled out under-reporting in 1983 and 1984 of serious injuries as an explanation, because this would have shown up in the Rutherford *et al.* (1985) analysis as an increase in numbers of rear seat passengers treated in hospitals; in fact, they found a slight decrease. We are reluctant to accept changes in driving behaviour as an explanation since these would be expected to lead to a corresponding increase in numbers seriously injured and there is no evidence of such an increase. For a similar reason the transfer of passengers from the front to the rear seats in response to the law is not a completely satisfactory explanation. We must therefore leave the sharp rise in the number of rear seat passengers killed relative to the number *KSI* as an unexplained mystery, at least until more evidence is available.

We now consider how to estimate the overall net effect of the seat belt law on numbers *KSI* and numbers killed. So far as front seat occupants are concerned, there is a clearly understood mechanism by which the wearing of a seat belt prevents or reduces the severity of injury and saves lives. Thus it seems reasonable to attribute changes in casualty rates for front seat occupants as due almost entirely to the effect of the law. However, for rear seat passengers, pedestrians and cyclists, the situation is more complicated. One has to try and assess whether changes in driving behaviour have taken place as a result of the law, and then attempt to disentangle the effect of any such changes from any other factors involved, such as possible changes in underlying trends, as in the case of pedestrians killed, and possible changes in passenger seating positions, as in the case of rear seat passengers killed. Direct evidence on either changes in driving behaviour or the other factors scarcely exists. Thus one feels less confident in attributing changes in casualty rates to the seat belt law than in the case of front seat occupants.

Overall, for the twenty three month period from February 1983 to December 1984, we estimate the reduction due to the law in numbers *KSI* for drivers and front seat passengers as 15,600. The estimated reduction in numbers killed is 879. We have included here figures for occupants of light goods vehicles, which are analysed in our Report though not in this paper. When casualties to rear seat passengers, pedestrians and cyclists are added, the estimated reduction in numbers *KSI* changes to 14,890 and the estimated reduction in numbers killed changes to 397. We conclude that, whether we concentrate on those directly affected or also include those indirectly affected, there have been substantial net reductions in numbers *KSI* and numbers killed due to the introduction of the seat belt law.

7.2. *Evidence on Changes in Driving Behaviour*

Some opponents of compulsory wearing of seat belts have argued that while casualties to front seat occupants will undoubtedly go down due to the introduction of the seat belt law, casualties to other road users, particularly rear seat passengers, pedestrians and cyclists, will increase to such an extent that the overall gains will be very substantially reduced or even cancelled out altogether. The reason given is that the protection conferred by wearing a seat belt will tend to make drivers feel safer and as a result some of them will change their driving behaviour in a way that will increase the risk of accidents. This theory is called the theory of risk compensation. The theory has been extensively discussed and there is a large literature on it. The leading exponent of the theory in this country is J.G.U. Adams and in Adams (1984, 1985) he discusses it using an interesting variety of evidence from a number of countries. An extensive review of work in the field is given by Ashton and others (1985). They concluded that "the available evidence indicates that risk compensation probably does not occur when drivers are compelled to wear seat belts".

Our own work was not directly concerned with testing the risk compensation hypothesis, and it will be apparent from the discussion in the previous sub-section that our evidence is somewhat contradictory. The figures for *KSI* casualties lead one to reject the risk compensation hypothesis. On the other hand, the killed figures do lend support to the hypothesis although, as we have noted, a careful look at the evidence indicates some reservations in arriving at such a conclusion.

7.3. *Statistical Methodology*

So far as methodology is concerned we have based our analysis on structural modelling instead of one the more conventional *ARIMA* system. Univariate *ARIMA* modelling is based on the idea that by differencing and other transformations a stationary series can be obtained. A parsimonious *ARMA* model is then selected by using statistical tools based mainly on the correlogram. Our experience is that truly stationary behaviour is often hard to achieve from real time series and one has to be content with approximations to stationarity that are not well defined. Moreover, the correlogram is rather treacherous as an instrument for model selection because of its high sampling variability for series of moderate length. Thus apart from straightforward cases such as those leading to the "airline model" the *ARIMA* model identification procedure is often quite hard to operate. Explanatory variables are incorporated into the *ARIMA* system by the transfer function technique. Even for a single explanatory variable this is not always feasible since the technique depends on the achievement of approximate stationarity by differencing and other transformations; see Harvey (1981, pp. 244–246). Moreover the basic identification tool is the cross-correlogram, which is even harder to interpret than the univariate correlogram and may be a very inefficient way of determining the form of a lag structure. The difficulties multiply as the number of explanatory variables increases.

The structural approach that we have adopted represents, we believe, a more direct and transparent technique for time series modelling. The basis idea is that one looks at the behaviour of an observed time series and where the series appears to contain trend and seasonal components one aims at modelling these directly. Allowance can be made for changes over time in the behaviour of these and other components as needed. Explanatory and intervention variables can be added in a direct manner. Because the models can be put in state-space form and thus can be handled computationally by the Kalman filter it is possible to accommodate a considerable amount of apparent complexity.

7.4. *Miscellaneous Points*

In our Report we mention at various points some difficulties that arose because of under-reporting of some accidents involving fatalities in the Metropolitan Police area. After submitting the Report we re-analysed data for Great Britain excluding this area but found that no differences of any consequence emerged. We have not therefore taken up space in this paper by discussing this re-analysis.

The data we used are available for research purposes from the Head of Computer Services, London School of Economics, Houghton Street, London WC2A 2AE. A nominal charge of £10 is made to cover handling charge and the cost of the floppy disc (available for IBM PC or BBC micro).

ACKNOWLEDGEMENTS

We are grateful to Mr E. J. Thompson, Director of Statistics at the Department of Transport, for inviting us to undertake this study, and to him and his colleagues, particularly Mr H. M. Dale and Mr P. J. Hathaway for their helpfulness at all stages. Our main debt is to our research assistant, Simon Peters, for carrying out the large amount of computing and data processing needed for this work. We also thank Richard Snell, Manuel Arellano and Javier Fernandez who played valuable supporting roles in this respect. Much of the methodology used was developed as part of the Economic and Social Research Council supported DEMEIC Econometrics programme.

REFERENCES

Adams, J. G. U. (1984) Smeed's law, seat belts and the emperor's new clothes. In *Human Behaviour and Traffic Safety* (C. Evans and R. C. Schwing, eds). Proceedings of International Symposium sponsored by General Motors. To appear.

——(1985) *Risk and Freedom: The Record of Road Safety Regulation.* Cardiff: Transport Publishing Projects.

Ashton, S. J., Thomas, P. D., Harms, P., Mackay, G. M. and Galer, M. D. (1985) The effects of mandatory seat belt use in Great Britain. Paper presented at the 10th International Conference on Experimental Safety Vehicles,Oxford, 1985.

Bhattacharyya, M. N. and Layton, A. P. (1974) Effectiveness of seat belt legislation on the Queensland road toll—an Australian case study in intervention analysis. *J. Amer. Statist. Ass.*, **74**, 596-603.

Bowman, K. O. and Shenton, L. R. (1975) Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, **62**, 243-250.

Box, G. E. P. and Jenkins, G. M. (1970) *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E. P. and Tiao, G. C. (1975) Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Ass.*, **70**, 70-79.

——(1976) Comparison of forecast with actuality. *Appl. Statist.*, **25**, 195-200.

Brown, R. L., Durbin, J. and Evans, J. M (1975) Techniques for testing the constancy of regression relationships over time (with Discussion). *J. R. Statist. Soc.* B, **37**, 149-192.

Durbin, J. (1984) Time series analysis. *J. R. Statist. Soc.* A, **147**, 161-173.

Durbin, J. and Harvey, A. C. (1985) The effects of seat belt legislation on road casualties in Great Britain: Report on Assessment of Statistical Evidence. Annex to *Compulsory Seat Belt Wearing Report by the Department of Transport*. London: HMSO.

Engle, R. F. (1978) Estimating structural models of seasonality. In *Seasonal Analysis of Economic Time Series* (A. Zellner, ed.), pp. 281–308. Washington, D.C.: Bureau of the Census.

Harrison, P. J. (1967) Exponential smoothing and short-term sales forecasting. *Manag. Sci.*, **13**, 821–842.

Harrison, P. J. and Stevens, C. F. (1976) Bayesian forecasting (with Discussion). *J. R. Statist. Soc.* B, **38**, 205–247.

Harvey, A. C. (1981) *The Econometric Analysis of Time Series*. Oxford: Philip Allen. New York: Wiley.

——(1984) A unified view of statistical forecasting procedures. *J. Forecasting*, **3**, 245–275.

——(1985a) Trends and cycles in macroeconomic time series. *J. Bus. Econ. Statist.*, **3**, 216–227.

——(1985b) Multivariate time series models, control groups and intervention analysis. CSE Econometrics Programme Discussion Paper No. A.53, London School of Economics.

Harvey, A. C. and Todd, P. (1983) Forecasting economic time series with structural and Box-Jenkins models: a case study (with Discussion). *J. Bus. Econ. Statist.*, **1**, 299–315.

Harvey, A. C. and Peters, S. (1984) Estimation procedures for structural time series models. LSE Econometrics Programme Discussion Paper No. A.44, London School of Economics.

Holt, C. C. (1957) Forecasting seasonals and trends by exponentially weighted moving averages. Carnegie Institute of Technology, Pittsburgh, Pennsylvania. ONR Research Memorandum No. 52.

Jarque, C. M. and Bera, A. K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.*, **6**, 255–259.

Kalman, R. E. (1960) A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Engning*, **82**, 35–45.

Kitagawa, G. (1981) A nonstationary time series model and its fitting by a recursive filter. *J. Time Ser. Anal.*, **2**, 103–116.

Ljung, J. and Caines, P. E. (1979) Asymptotic normality and prediction error estimators for approximate system models. *Stochastics*, **3**, 29–46.

Maravall, A. (1985) On structural time series models and the characterization of components. *J. Bus Econ. Statist.*, **3**, 350–355.

Muth, J. F. (1960) Optimal properties of exponentially weighted forecasts. *J. Amer. Statist. Ass.*, **55**, 299–305.

Nelson, C. and Kang, H. (1984) Pitfalls in the use of time as an explanatory variable. *J. Bus. Econ. Statist.*, **2**, 73–82.

Nerlove, M. and Wage, S. (1964) On the optimality of adaptive forecasting. *Manag. Sci.*, **10**, 2, 207–229.

Nerlove, M., Grether, D. M., and Carvalho, J. L. (1979) *Analysis of Economic Time Series*. New York: Academic Press.

Pierce, D. A. (1978) Seasonal adjustment when both deterministic and stochastic seasonality are present. In *Seasonal Analysis of Economic Time Series* (A. Zellner, ed.), pp. 242–269. Washington, D.C.: Bureau of the Census.

Road Accidents in Great Britain, 1983 (1984) London: HMSO.

Rutherford, W. H., Greenfield, A. A., Hayes, H. R. M. and Nelson, J. K. (1985) *The Medical Effects of Seat Belt Legislation in the United Kingdom*. Research Report No. 13, Office of the Chief Scientist, DHSS. London: HMSO.

Schweppe, F. (1965) Evaluation of likelihood functions for Gaussian signals. *IEEE Trans. Inform. Theory*, **11**, 61–70.

Scott, P. P. and Willis, P. A. (1985) Road casualties in Great Britain—the first year with seat belt legislation. TRRL Research Report 9, Department of Transport.

Theil, H. and Wage, S. (1964) Some observations on adaptive forecasting. *Manag. Sci.*, **10**, 198–206.

Tiao, G. C. (1972) Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, **59**, 525–531.

Wilde, G. J. S. (1982) The theory of risk homeostasis: implications for safety and health. *Risk Analysis*, **2**, 209–225.

Winters, P. R. (1960) Forecasting sales by exponentially weighted moving averages. *Manag. Sci.*, **6**, 324–342.

DISCUSSION OF THE PAPER BY PROFESSORS HARVEY AND DURBIN

Dr C. Chatfield (University of Bath): I welcome today's paper which is important from both a methodological and a subject-matter point of view. The Society has the chance to discuss a topic of wide public interest, and also to compare the merits of structural and *ARIMA* modelling with other time-series approaches.

The class of structural models looks suitable for many problems. I recommend reading Harvey (1984), but see also the cautionary remarks in the discussion of Harvey and Todd (1983). The state-space formulation is reminiscent of the dynamic linear model of Harrison and Stevens (1976), but the Kalman filter is used in a non-Bayesian way which may increase its appeal.

I agree with much of what is said about ARIMA modelling, in particular that it is often undesirable to eliminate trend and seasonal effects by differencing (e.g. Chatfield, 1978, p. 227). However, *ARIMA* models should continue to be useful for some series which are not dominated by trend and seasonal variation. One good aspect of the Box-Jenkins approach is the emphasis on model-formulation, and perhaps more could be said here on how to *choose* an appropriate structural model. In particular the authors do not explain why they take logs of the data. Is it because they think the seasonal effect and/or the error variance is multiplicative? Wherever possible I prefer to model the raw data. The seasonal effect does appear to be additive (see below) while changes in error variance look generally "small".

The authors concentrate on what they call the basic structural model (*BSM*), which is closely related to the Holt-Winters (HW) forecasting approach (e.g. Chatfield, 1978). The latter is easy to use and to understand and has performed well in recent forecasting competitions (e.g. Makridakis *et al.*, 1982). However the structural approach is based on a proper stochastic model. This means for example that prediction *MSE*'s can be calculated. In addition the structural approach can more easily handle irregularities in the data and be extended to incorporate explanatory variables. However, despite these advantages, I am rather uneasy about its early use on a topic of public importance. The approach is not yet well-tried, is rather complicated for the non-statistician to comprehend, and in any case I suggest that many of the findings could be obtained with a simpler descriptive approach. I think the latter should always be tried anyway, both to get a "feel" for the data, and as a yardstick, even if it is found that it needs to be backed up by a more complicated model-based approach. No doubt the authors have tried a simpler approach (see Galley 2), but this is not always evident in the paper.

I therefore carried out an initial examination of the data (or IDA—see Chatfield, 1985) which here consists mainly of looking at appropriate averages and at time plots of the different series. As it is hard to say which categories of road user are, directly or indirectly, affected by seat-belt legislation, I began by looking at 'grand total' figures as shown in Table D1. I have extended the table to 1985 with provisional figures. The decline in total casualties since 1982 is relatively small, but the reductions in killed (*K*) and seriously injured (*SI*) appear substantial and consistent. I think the authors are right to concentrate on the latter. The roughly 10% drop in people KSI provides a powerful justification for the retention of seat-belt legislation, although one cannot yet rule out all other possible explanations.

When we look at the constituent series in more detail, the authors' Fig. 1 shows a 'clear' reduction in car drivers *KSI* in 1983/4 and many people will wonder if it is really necessary to show that it is "significant". Other time plots show for example little change in pedestrians killed, but perhaps some increase in cyclists KSI. The power of IDA may be further demonstrated by considering rear seat passengers killed, which the authors are surprised to find have increased significantly. Annual totals killed for all car occupants for 1978–84 are shown in Table D2. The average yearly increase for rear-set passengers is around 20% (compared with the 27% found by the authors). Two obvious explanations are that more passengers are travelling in rear seats (to avoid wearing seat belts) and that the police are classifying all unbelted passengers as 'rear seat'. As evidence, I note that the reduction in front-seat passengers killed is surprisingly larger than for car drivers. This suggests looking at all passengers combined and we then find that there is indeed an overall improvement, though understandably not as large as for car drivers. This largely resolves this apparent puzzle.

To assess the seasonal variation, and hence more accurately assess the improvement since February 1983, I next ran the monthly car drivers *KSI* data through my Holt-Winters (HW) programs. A slightly better fit is obtained with an additive, rather than a multiplicative, seasonal model. The forecasts from January 1983, suggest a 22% reduction—agreeing well with the authors' results. The HW results are easy to obtain and can readily be found for the other series as well. Although the HW method does not provide exact confidence intervals, an upper bound can be obtained from the 1-step-ahead error variance which shows that the decrease for car-drivers *KSI* is "significant".

TABLE D1
Total road casualties in Great Britain

Year	*Total casualties*	*Total killed*	*Total Seriously Injured*
1980	329,000	6,010	79,000
1981	325,000	5,850	78,000
1982	334,000	5,930	80,000
Average (before)	329,000	5,930	79,000
1983	309,000	5,440	71,000
1984	324,000	5,600	73,000
1985	320,000	5,200	71,700
Average (after)	317,700	5,410	72,000
Change in yearly average	−11,300	−520	−7,000

Source: Monthly Digest of Statistics (with a little rounding)

TABLE D2
Car occupants killed, yearly totals

Year	*Car drivers*	*Front-seat passengers*	*Rear-seat passengers*	*All passengers*
1978	1525	687	293	980
1979	1479	632	280	912
1980	1339	637	263	900
1981	1346	613	304	917
1982	1472	658	297	955
Average (before)	1432	645	287	932
1983	1198	480	321	801
1984	1228	539	372	911
Average (after)	1213	509	346	856
Change in yearly average	−15% or −219 lives	−21% or −136 lives	+20% or +59 lives	−8% or −77 lives

ıtial side-benefit of the *BSM* may be to provide approximate prediction interval results for at other purposes do we need the BSM here? For some series, it is not clear from the IDA or the use of HW if the changes are "significant". However, it is the overall change which is of prime importance and this is arguably clear. In any case it is hard to say what is meant by "significance" in the context of complete population figures. Where we *do* need the *BSM* is to assess the form of the intervention and the effects of explanatory variables. As to the latter the results look disappointing. Section 4.3 shows that R^2 only increases from 0.76 to 0.78. Although "significant", the explanatory variables make little practical difference. On grounds of simplicity, I therefore suggest excluding these explanatory variables. It is a pity that time did not allow the consideration of other variables. Nevertheless it seems that the BSM provides a powerful new way of investigating explanatory variables.

It is clearly a matter of opinion as to how much of the analysis today needs to be based on the BSM. Some statisticians seem to feel unhappy using descriptive methods, but I believe that simple methods are often superior and that we need to be more clear as to why and when we need to use more complicated methods. In this case, I suspect that any reasonable method (even including Box-Jenkins!) will spot the substantial reduction in numbers KSI, so that this case study of structural modelling may be regarded as only partially convincing. So let me conclude by stressing two points; first that my analysis points in the same direction as that of the authors; second that I regard the class of structural models as a potentially valuable addition to the time-series toolkit. It therefore gives me much pleasure to propose the vote of thanks.

Dr G. Tunnicliffe Wilson (University of Lancaster): First, may I say that this study, by its thoroughness, does a great service not only to the analysis of transport statistics, but also to the wider reputation of statistics. The authors have pursued many avenues in detail and with care, to check their conclusions. That is a great credit to them.

Concerning the debate about the relative merits of structural and *ARIMA* models, I would hope that these might give consistent results, with any minor differences being resolvable. For this study I have re-estimated the intervention coefficients using the 'Airline' *ARIMA* model, and find that for Car Drivers *KSI*, and for Pedestrians Killed by Cars, exact agreement to the three d.p. published in the paper. For Rear Seat Passengers Killed, the estimate was .238 compared with .237. The only slight difference is that evidence of a constant seasonal component, corresponding to a seasonal *MA* parameter $\Theta = 1.0$, is only confirmed in one case, possibly because the *ARIMA* parameters were estimated on the full data set. There are small disagreements in estimated standard errors. Might I suggest the publication of likelihood ratios as test statistics, as possibly providing closer agreement.

On the wider issues of this debate, the claim that structural models 'have a direct interpretation' must be tempered by the fact that several structures, compatible with each other and with the data, may be possible. The particular choice of model may be for convenience in parametrisation, and also subjective appeal. This is most apparent in the structure of the seasonal aspects of the model, several forms having been advocated in the literature. On the other hand, *ARIMA* models acknowledge the fact that series arise from complex systems, and seek empirically to capture the dynamics, with no pretence that they can represent the true structure. Nevertheless, the 'Airline' model has a constraint in its structure, imposed by the multiplicative differencing and moving average operators. Thus *both* for the *ARIMA* model, and for what might better be termed the Independent Components model of this paper, there is a balance in the choice of structure between subjective constraints and empirical generality.

Concerning the authors' criticism of differencing, may I point out that *ARIMA* models are called Integrated, not differenced Models, and the operator ∇ has the same rôle as in their structural models. Their criticism may be valid, of the use of differencing in the data analysis process. I would agree that this requires an understanding, which the authors have displayed, and which most analysts can acquire.

Turning now to the search for interventions, I suggest it would be useful to have some way of displaying where step jumps have occurred throughout the series. For example the Harrison-Stevens mixed process *DLM* detects such jumps with no prior specification of their location. A linear tool of use here, has been suggested by Professor G. C. Tiao. This is to look for outliers in the differenced series $w_t = \nabla x_t$, by graphing $z_t = w_t - \hat{w}_t$ where $\hat{w}_t$ is the value interpolated for w_t from the rest of the data, using the *ARIMA* model with no intervention variable. In fact

$$z_t \propto (1-\theta B^{-1})^{-1}(1-\Theta B^{-12})^{-1}(1-B^{-12})e_t,$$

where e_t is the residual series from the *ARIMA* model for x_t.

For the series of Pedestrians Killed by cars, z_t is graphed in Fig. D1 with $2SE$ limits. The evidence from this is that significant positive steps took place in November and December 1982, before the legislation. However, the corresponding graph (not shown) for Rear Seat Passengers Killed clearly points to February 1983.

Finally, I wish to take up the point on sensitivity analysis in 5.4. In terms of the structural model, assessment of any intervention jump depends on the general fluctuations in level of the series, i.e. the size

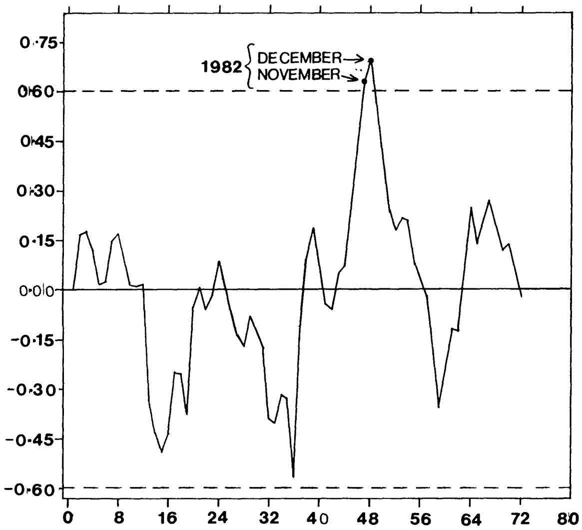

Fig. D1. Estimated (scaled) jumps in pedestrians killed by cars, with 2 *SE* limits, graphed against months between January 1979 and December 1984.

of its Random Walk component as measured by the ratio $\sigma_\eta^2/\sigma_\varepsilon^2 = R$ say. Taking the seasonality as fixed, this is related to the moving average parameter by $R = (1 - \theta)^2/\theta$. For pedestrians killed by cars, R is estimated as 0 (thus $\theta = 1$). I have extended the analysis of 5.4 by graphing in Fig. D2, not only the estimated intervention coefficient and its *SE* for θ in the range (0.5, 1.0), but also the likelihood associated with θ. Values of θ well away from 1.0 are quite plausible, and for these the coefficient is 'not significant'. A Bayesian interpretation would not result in a particularly large probability that the coefficient was positive.

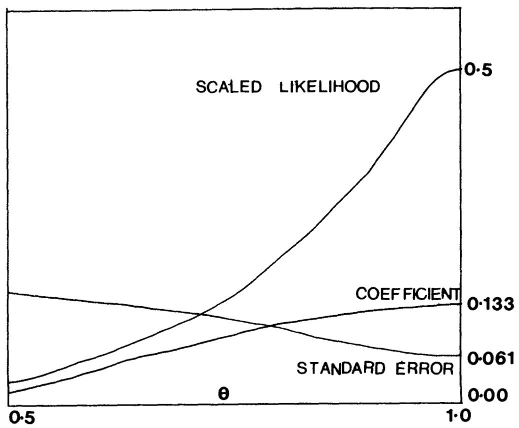

Fig. D2. Estimated coefficient and its standard error for the intervention in pedestrians killed by cars, graphed for values of the moving average parameter θ between 0.5 and 1.0, together with the marginal likelihood of θ.

Within this paper I have found much to impress and stimulate me, and I am very pleased to second the vote of thanks.

The vote of thanks was passed by acclamation.

Mr E. J. Thompson (Dept of Transport): May I first record the Department's thanks to Professors Harvey and Durbin for agreeing to undertake their study at short notice, for producing the results within the agreed timetable, and for providing such a clear report.

Working with them has been a very stimulating experience for the Department's statisticians. They were able to provide a fresh, dispassionate look at the data we work with constantly, bringing to their analysis technical expertise of an order not usually found within the Government Statistical Service. We have seen a notable example of effective cooperation between the academic world and the GSS — two important parts of the community represented in the RSS.

This issue provides an excellent example of real statistics — with all their inevitable imperfections — affecting important decisions in the real world. The analyses incorporated in the Department's report — that by Durbin and Harvey being accompanied by reports on research by the Transport and Road Research Laboratory, William Rutherford's analysis of hospital admissions data, and the much simpler statistical analyses by the Department's statisticians — clearly convinced many waverers (myself included!) that seat belt *legislation* had saved many lives, and reduced the severity of many injuries, thanks to the remarkable and unpredictably high wearing rates which followed it. It is encouraging to note that the further independent analyses reported in today's discussion support that conclusion.

Preliminary data for 1985 confirm the step down in the levels of deaths plus serious injuries after 31 January 1983. There is no sign of it wearing off. My only criticism of the paper is that it gave no projections for 1985 to compare with the actuals to be published in four weeks time!

Professor M. B. Priestley (University of Manchester Institute of Science and Technology): It's nice to see time series analysis in action, particularly when it is applied to a real problem of considerable importance rather than serving merely as an illustration of an author's new technique. Tonight's paper presents a high quality and detailed piece of time series analysis concerning data of considerable importance, and as such it represents a very valuable addition to the literature.

The authors's *BSM* model (2.3a, b) clearly provides a very good description of the data, and this is confirmed by their impressive array of diagnostic aids. However, I am not clear as to the basic difference which the authors see between the *BMS* and the *ARIMA* class of models. They say that they are not happy with the assumption that differenced series can be assumed stationary, but this is precisely what they are assuming in setting down the model (2.3a, b). In saying that the ε_t, η_t, and ζ_t are stationary white noise processes, they are in effect saying that the second difference of the y_t are stationary. Moreover, with $\sigma_w^2 = 0$ the model (2.4a, b, c) for the seasonal component reduces to a simple Fourier series, and this again is entirely equivalent to saying that the seasonal component is described by the difference equation $(1 - L^{12})\gamma_t = 0$. It is certainly true that structural models enjoy considerable flexibility but, as Dr Tunnicliffe-Wilson has pointed out, the choice of a structural model is essentially a subjective matter, and it is easy to envisage a wider class of structural models than those described in the paper. (One restriction of the model (2.3a, b) is that we cannot generate *ARIMA* models with general autoregressive operators; the *AR* operator here consists purely of difference and seasonal difference operators.)

One of the difficulties that arise in comparing classes of models is the interpretation of standard terminology. In this context it should be noted that the terms "*ARIMA*" and "Box-Jenkins methodology" are by no means synonymous. Many analysts identify and fit *ARIMA* models using techniques quite different from those originally used by Box and Jenkins and there are now much more sophisticated methods for identifying the form of an *ARMA* model than that based on a scrutiny of the sample autocorrelation function. (I refer e.g. to the various order determination criteria which are in wide spread use for model selection.) Similarly, the terms "Kalman Filter" and "state-space models" are by no means synonymous. The Kalman Filter is simply an algorithm which is *applied* to state space models: it does not in itself constitute a "model". Although it is true that Kalman first proposed his algorithm as far back as 1960, its application to time series analysis could hardly have been foreseen at that stage since the fundamental connection between state space models and *ARMA* models was not fully clarified until the early 1970's (see, e.g. Akaike, 1974a, b).

The general conclusions derived from the authors' use of intervention analysis seem quite indisputable, but it would have been interesting if they had moved the "change point" further back in time and then introduced some "dynamics" for the effect of the intervention variable. The fact that the event of interest occurs only once does not prevent the estimation of a more general "transfer function" approach and this would have obviated the inclusion of the somewhat arbitrary value of w_t set for January 1983.

As regards the use of explanatory variables, it is rather surprising that the petrol price index seems to have a rather small effect, since a purely visual inspection of the *KSI* series would indicate something akin to a step change occurring round about January 1974. This period occurs only shortly after the sudden increase of petrol prices following the oil crisis, and it would be interesting to see whether the authors' approach confirms an "intervention effect" at the beginning of the year 1974.

In conclusion, may I congratulate the authors once again on presenting us with a really excellent and extremely interesting piece of time series analysis.

Professor J. P. Burman (University of Kent at Canterbury): I too welcome this paper, as being in the best traditions of the Society: a topic of great public interest combined with up to date statistical analysis.

My experience of Time Series modelling has been entirely of the *ARIMA* type. There are advantages for the state-space method in retaining the intuitive link with trend and seasonality. But I think the authors somewhat exaggerate the difficulties of *ARIMA* model identification, and minimise those of the component models: one still needs the correlograms for selecting the appropriate trend and irregular component models. Fig. 3 tells us that $\Delta_{12} y_t$ is stationary, and Fig. 2 confirms that we do not need both Δ and Δ_{12}, since r_1 is -0.55, suggestion over-differencing (see top of p. 195). To find an *ARIMA* global model, I should start with either the airline model or (1, 0, 1) with a fixed seasonal pattern. In the first case, my estimation program might find Θ below -0.95, and replace the stochastic by deterministic seasonality. One would probably arrive at the same final model in either case.

But of course the usual factorisable *ARIMA* family does not contain the *BSM*: the former has a simple *ACF*; the *BSM* has a complex one. Is it not possible to compare the fit of the two models, using R^2 and the Box-Ljung Q test?

I have for long been puzzled about how the Kalman filter can provide Maximum Likelihood rather than Constrained Least Squares estimates: how are the start values of the ε_t, η_t, ζ_t, and especially the 12 ω_t obtained? I have looked at the Schweppé reference, but could not find any explanation of this.

The rises in the numbers killed since 1982 shown in Table 2 for certain categories demand some explanation. For rear-seat passengers R^2 is only 0.45, which may mean that important variables have been omitted. The authors note that they have ignored the effect of differing numbers of days of the week in a month: is it not likely that there are far more rear-seat passengers at week-ends? The same question arises, less obviously, for cyclists. But it is *a priori* unlikely that the omission of short-term calendar effects would eliminate the under-prediction of the models which omit an intervention variable.

I am very glad that the authors have obtained permission to release the data, and I hope to experiment with them myself.

Mr G. J. A. Stern (I.C.L.): I believe that the authors' work was commissioned by the Department of Transport in significant part in response to Dr John Adams's work, and Dr Adams ought to be given at least as long as the proposer and seconder to comment.

In his "Histoire de Charles XII", Voltaire in his preface opens with the words "Incredulity, said Aristotle, is the foundation of all wisdom", and he goes on to say that he not only disbelieves accounts of prodigies but also stories which, though not impossible, lack all probability. Commenting on Plutarch's well-known account of Caesar swimming, fully armed, holding papers over his head to keep them dry, Voltaire says "Do not believe a single word of this story which Plutarch tells you . . . if you jump in the sea with papers in your hand, you get them wet." (Voltaire, 1750) Dr Adams proposes that if you feel safer in your car because of a seat-belt (or other reasons) you are likely to drive more dangerously at least as far as other road users are concerned. He supports this (Adams 1985) with a mass of data from many countries which shows that road deaths did not decrease after seat belt laws were introduced any more than they decreased in countries without such a law. It seems to me that the assertion made by the Department of Transport that a safer-feeling car does not cause one to drive more dangerously falls well within the category of stories lacking all probability which Voltaire rejects. I feel that statistically the authors have written an interesting and useful paper, and instinctively I tend to prefer their approach to Box/Jenkins. I commend them too for agreeing that the road death figures do give some support to Dr Adams because they show less drivers but more back-seat passengers, cyclists and pedestrians killed after the seat-belt law. The authors ought to have concentrated on the deaths data because serious injury data is very badly reported and unreported—Adams (1985) cites a T.R.R.L. study showing 59% non-reporting of serious accidents to cyclists, and 21% non-reporting to road users overall, while deaths were fully reported. The authors ought to have been commissioned by the Transport Department to look at overseas data. As regards British data, the Department ought to have asked for the effect of the new breathalizer and stricter drink-driving enforcement to be taken into account. On this point, Lord Caithness, speaking for the Department, wrongly claimed that all factors, "including drinking" had been taken account of. (Hansard, 1986). I fear that complex analysis such as this, by excluding most people from the debate, can mask fudging by the authorities. A series of graphs, like Dr Adams's, showing deaths before and after seat belt laws all over the world, with added comments about drink-drive enforcement etc, could have been more useful to the average M.P and peer—the fault here is the Department's. I would propose "the GT10 test" for statistical papers which will influence politicians: given that politicians are people of many cares working in an establishment with twenty to thirty ever-open bars, we should ask "How clear is the paper after the tenth G & T?" On this test the paper, admirable in many ways, fails.

Mr H. M. Dale (Dept of Transport): I too would like to congratulate Professors Durbin and Harvey for their contribution to the public debate on this subject. Their report to the Department of Transport was a model of clarity; it was not afraid to point out problems, but at the same time made clear where robust conclusions could be drawn.

One puzzle remaining at the end of the study was the apparent discrepancy between the effects on fatal and serious injuries. The problem with operating on fatality data is that the numbers, when disaggregated, are rather small. The time-series are therefore subject to greater variability and we may have witnessed unusually high fatality figures in 1983 and especially in 1984. I would therefore place more reliance on the figures for killed and seriously injured combined to estimate the effect of the regulations.

Data now available for January to September 1985 confirm the volatility of the fatal casualty series. The indications are that road deaths in 1985 will be nearly 10 per cent lower than in 1984, while the decline in serious injuries may be only about 2 per cent. My Table D3 updates Table 1 of Durbin and Harvey's paper.

TABLE D3

Numbers killed and seriously injured: 1982, 1984, 1985

	Killed and Seriously injured			*Killed*		
	1982	1984	1985(*F*)	1982	1984	1985(*F*)
Car drivers	19,460	16,530	16,700	1,472	1,237	1,260
Car front seat passengers	9,458	7,100	7,000	658	544	470
Car rear seat passengers	4,706	5,092	5,000	297	374	350
Pedestrians	18,963	19,461	19,400	1,869	1,868	1,680
Cyclists	5,967	6,595	5,600	294	345	290

F indicates forecast based on partial data

Mr John G. U. Adams (Geography Department, University College London): Professors Harvey and Durbin conclude that when all categories of road user are considered the net reduction in fatalities attributable to the seat belt law is about 200 per year (from 5934 in 1982). There are four reasons for supposing that they have over-estimated the beneficial effect of the law.

Their estimate of the increase in pedestrian fatalities (8%) and cyclist fatalities (13%) is based on the modelling of *all* such fatalities. In their Report, but not their paper, they estimate the increase in numbers of pedestrians and cyclists killed by cars and vans at 13% and 40% respectively. The number of pedestrians and cyclists killed by HGVs and PSVs decreased.

The serious injury data and the fatality data tell different stories. Durbin and Harvey prefer the serious injury data because the numbers are much larger. However the fatality data are much more accurate. For pedestrians and cyclists (categories not covered by the Rutherford study) injuries are grossly under-reported. The definition of serious injuries is acknowledged by the BMA to be so broad as to be virtually useless.

In 1982 the percentage of dead drivers with blood alcohol levels above the legal limit increased from 31% to 36%. All of the increase in fatalities in 1982 occurred during the "drink-drive hours" (2200–0400). Thus the increase in fatalities in 1982, when some of their models show signs of breaking down, could have been an alcohol effect. In 1983 the percentage of over-the-limit dead drivers fell back to 31%. The

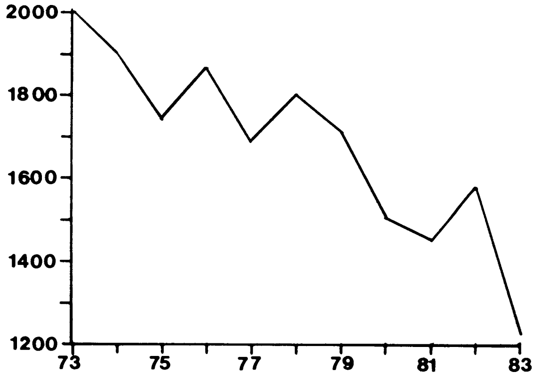

Fig. D3. Road accident deaths, GB, between hours 22.00 and 03.59.

reduction in fatalities in 1983 during the "drink drive hours" was 23%, and in all other hours only 3%. Durbin and Harvey attribute *all* of the decrease in front seat car and van fatalities in 1983 and 1984 to the belt law. Their model includes no alcohol variables, and they allow the reduction in drunken driving in 1983 credit for saving not a single life.

Durbin and Harvey made no attempt to place the British evidence in the context of evidence from other countries which have passed belt laws. Because of the possible effects of innumerable confounding variables, the results from one country in isolation may be inconclusive. But there is no clear evidence from any country of a net beneficial effect attributable to a belt law (Adams, 1985).

If one trusts the accurate numbers (fatalities) rather than the large numbers (serious injuries), if one isolates the pedestrians and cyclists killed by cars and vans, if one allows the reduction in drunken driving a reasonable share of the credit for the decrease in fatalities in 1983, and if one considers the evidence from other countries, the balance of Durbin and Harvey's evidence tilts strongly in favour of the conclusion that there has been no net life-saving benefit attributable to the belt law—only a shift in the burden of risk from the best protected to the most vulnerable road users.

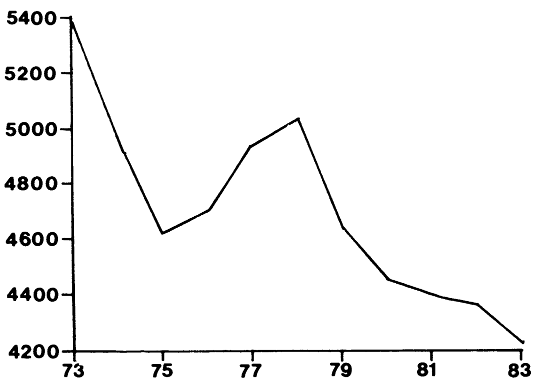

Fig. D4. Road accident deaths, GB, between hours 04.00 and 21.59.

Mr F. West-Oram (The Pedestrians Association): We are not opponents of compulsory seat-belt wearing. We support any measure for saving life on the roads. However, we are naturally disturbed by any evidence of rising pedestrian casualties.

Durbin and Harvey (1985) reported a significant 14 per cent increase in pedestrians killed *in two-party collisions involving cars or light vans* (i.e. with belted drivers) following the introduction of the seat-belt law. There was a non-significant 8 percent increase from all collisions including those involving unbelted drivers. The 14 per cent rise was equivalent to an estimated increase of 131 pedestrian deaths in 1983 and 132 in 1984.

The authors reported no significant change in pedestrians killed and seriously injured. However, they had no serious-injury hospital data for pedestrians. Hobbs, Grattan and Hobbs of the Transport and Road Research Laboratory (1979) showed that this information does not consistently find its way into the official statistics noting 18 per cent under-reporting in their investigation.

Direct comparison of the "two-party" pedestrian casualties for 1984 (belted) with 1982 (unbelted) shows 8.9 per cent increase for deaths and 7.5 per cent for serious injuries. This comparison is simplistic, but does raise questions about the seriously-injured data available to the authors.

They were refused information by the insurance authorities on a possible disproportionate increase in damage-only collisions (in which unreported pedestrian casualties might have occurred). Reports of increased damage-only claims have appeared in the press.

The authors were apparently not given details of increased drink-drive testing in 1983 and 1984, which was associated with about 20 per cent reduction in car-user deaths in the "drink-drive" hours 22.00–04.00, but only 2 per cent in the remainder of the day; and with a lower proportion of drivers found killed with excessive blood alcohol content—from 36 per cent in 1982 to 28 per cent in 1984. These factors could cause the reductions claimed for seat-belt use to be considerably offset.

We feel there is strong evidence supporting risk-compensation. However, the main message is clearly the need for much higher priority on reducing the actual collisions, rather than on passive measures that in this case reduced overall deaths by no more than one twentieth and benefitted one group of road users only.

Mr D. H. Jones (Middlesex Polytechnic): I would like to bring to the Society's attention a study which considers the "transfer of passengers" from front to rear seats—a subject mentioned briefly in Harvey and Durbin's interesting paper.

Researchers at Middlesex Polytechnic and the TRRL carried out two surveys on the ages, seating position and restraint usage of children in cars before (November 1982) and after (June 1983) legislation on seat belt use. Fifteen "site-days", chosen to be representative of a variety of national road conditions, were observed with data on 4967 children in 3438 cars and 7532 children in 4904 cars being recorded in the before and after studies respectively (Lowne *et al.*, 1984).

In analysing the data, differences were observed between the two surveys in the age distribution of the children, the numbers of children per car and the proportions of cars carrying front seat adult passengers. "Before" data were adjusted to eliminate the effects of these differences, enabling a direct comparison of the location and restraint usage of the children in both studies to be made.

Overall, there was no significant change in restraint usage rates. However, in the after study, more babies (less than 1 year old) were located in the rear, where more were restrained; more small children (1–4) were seated in the rear where *less* were restrained; more medium children (5–9) were in front seats with greater restraint usage and large children (10–14) had increased restraint usage without change in location.

If those who are unrestrained are considered to be "at risk", the proportion at risk fell for babies, *rose* for small children and was unchanged for medium and large children. Evidence observed later from national statistics on children mortally and seriously injured showed similar patterns for these age groups.

This study assessed the effect of "migration" after legislation on the population of children "at risk". Those who propound the theory of risk compensation should take into account the effect of such shifts in population when comparing pre and post legislation injury statistics. In retrospect, it is unfortunate that data on the seating position and restraint usage of adult passengers in all cars were not collected in our surveys.

Mr P. P. Scott (Dept of Transport): This paper is a valuable contribution to the analysis of road accident data. Considering the large volume of data on accidents which has been gathered over the years, comparatively little modelling of that data has been attempted. Researchers in the subject should find the methods reported here potentially useful.

I would like to ask two questions. Firstly, is it the case that the parameters representing variability in trend and seasonality will "mop up" variation due to identifiable factors which have been excluded from the model? An example of these is the fuel crisis period of December 1973 to Spring 1974; this will not (as the authors suggest in their Section 4.3) be represented particularly well by the petrol price variable, since *that* reached its peak values somewhat later.

Secondly, in Section 5.1 it is stated that observations several time intervals after the imposition of the seat belt law will have little influence on the estimate of λ, the parameter measuring the effect of the law. Doesn't the same apply to observations well *before* the law was introduced? Certainly a long series is needed in order to estimate the variances in the model, but if the objective is just to estimate the effects of the law then simpler models fitted to shorter series might be adequate.

I have one further comment. At the beginning of section 4.3 the authors mention that Scott and Willis (1985) assumed a simple regression structure. This is true, but the assumption was made in the light of previous analyses of accident series through the 1970s which had shown that little was gained (except a lot of hard work) by fitting ARIMA models instead of "ordinary" regression models, even in cases where residual diagnostics indicated that the regression fit was rather unsatisfactory. Of course, structural models were not investigated.

Professor A. S. C. Ehrenberg (London Business School): This really is a dreadfully vapid paper—full only of mental masturbation: The danger is that younger statisticians often think such performances sophisticated and ape them.

Why did the authors not simply report that all their "time-series" mumbo-jumbo made no difference? Their *Conclusions* say "We found reductions of 20-odd% for car drivers, 30% for front seat passengers, and so on". What do they mean "We found"? Anyone with enough courage like Dr Chatfield to do simple percentages on the raw data as in Table 1 would have "found" much the same: Car drivers down by about 20%, front seat passengers by about 30%, and so on. "We found!"

Or again: "The fact remains that we find the large increase in rear seat passengers killed hard to explain". That is not surprising since in the main paper the authors never even *tried* to explain it.

It is not a question of being sophisticated, rigorous, or mathematical on the one hand, and a bit simplistic with simple percentages on the other. Instead, the distinction is, as always in my view, between:

(A) Applying supposedly sophisticated technical procedures or models to data one has not yet understood,

(B) First finding some simple patterns in one's data, then seeing if they generalise to different conditions (e.g. different months and years here, different parts of the country; very experienced drivers and greenhorns; town, motor-way, and country traffic; high speed and low speed; rush hours and other times; the number of passengers even, and other countries), and only then bothering perhaps to look for models of what by then one already knows.

The latter approach certainly works with much more complex sequences of observations than the simplistic seat-belt data here, including panel-type measurements on the same individual items or people over time. To give some examples:

— One involving thousands of purchase sequences over time was presented here at the Society a couple of years ago (Goodhardt *et al.* 1984).

Or more recently:

— Two papers last week on people's attitudes over time (Lievesley and Waterton 1986, Barnard *et al.* 1986).

— An ESRC research proposal this week to collected dozens of experimental sets of consecutive readings over 6 months with literally hundreds of deliberate "interventions" (Ehrenberg and England, 1986).

— A report last month of people's choice of different TV programmes over time (Ehrenberg, 1986).

— A U.S. study a few weeks before of people's choice of different episodes of the same TV programmes over time (Ehrenberg and Wakshlag, 1986).

I could go on.

All these are examples of "time-series". Yet I do not see that any of this paper's techniques could ever be relevant. After all, they don't even seem to have been relevant with seat-belts.

Professor R. L. Smith (University of Surrey): I found this paper to be an impressive example of structural time series modelling. It is certainly a positive feature of such models that additional variables, such as explanatory variables and the intervention term, can be brought into the model without any change in the basic mathematical structure.

In much of the discussion, it seems that λ, the coefficient of the intervention term, is a critical parameter: it cannot be estimated consistently, but many of the conclusions depend on it. A plot of the profile likelihood of λ (i.e. the likelihood function maximised with respect to all the other parameters) may well be useful in assessing the information about λ in the data. Alternatively, since the profile likelihood often has poor properties when there are many nuisance parameters, one could switch to a Bayesian approach. The structure of the Kalman filter lends itself well to Bayesian analysis, and many of the required integrations could be carried out analytically. It should be possible to calculate the marginal posterior density of λ without a great deal more numerical computation than is already in the paper. In my opinion, this would be more informative than simply quoting an estimate and its standard error.

Professor Violet Cane (Statistical Laboratory, Cambridge): I must congratulate the authors on a beautiful piece of work at macro level — that is, one national data — but I feel that some of the anomalous results which they have pointed out need discussion on the micro level. ("Micro" for me means the city and county of Cambridge where I can get data rather easily.)

First, the curious case of the rear seat passengers — that is, the smallish increase of severely injured compared with killed. It is possible that we have run out of rear seat passengers. What we need is some information on car occupancy. Visual inspection of commuter traffic in Cambridge and Manchester suggests that about 90 per cent of cars contain only the driver. I accept that this is not true through the whole day. However, the data supplied in the paper suggests that only 50 per cent of cars have even a front seat passenger. It could be that there are not, as it were, people in the rear seats, who otherwise would have had minor injuries, to move up into the severely injured category.

Secondly, as far as the difference between the front and rear seats is concerned, the Cambridgeshire police say that this is a buffer effect. Previous to the seat belt regulations the front seat passengers provided a flexible buffer for those in the rear, whereas now they form an inflexible buffer.

Thirdly, the low importance of the traffic index for car drivers and passengers may be an effect of looking at national figures. We have very detailed data in Cambridge on all accidents and for individual radial roads the number of accidents is in general proportional to the amount of traffic. The exceptions to this are roads which have high numbers of shoppers whose presence changes the whole relationship. A mixture of very different roads taken over the whole country may well produce some odd effects.

Similarly, the less good fit of the model for pedestrians and cyclists could be due to the fact that their conflicts with cars mainly occur in towns. Increased pedestrianization of city centres and provision of urban cycleways may well have reduced the number of persons at risk.

In general I am worried about the extreme difficulty of explanation of some cases. For instance, in Cambridge the total number of accidents increased by 30 per cent between 1981 and 1982. Nobody has ever been able to find out *why*, although data have been analysed in every possible way.

Could higher car speeds be a cause of many of these troubles? I am not talking about the effect of petrol price but of a gradual increase each year. Injuries would be more severe and also cars would pass more cyclists on any given road. (The number of cyclists passed on average would be proportional to relative speed over car speed.)

Finally, higher speeds make prediction of movements on the road much more difficult. In fact, road users have to do a lot of calculation in order to predict where traffic they can see now will be two seconds later. Some people seem to suffer fromwhat I can only call "dyskinesia", especially the old, the young and the drunk. There may be people who are not capable of such calculations, and there are others who will not both through selfishness. Perhaps a suitable psychological test should be devised.

Mr Michael J. Stewart (Abbey Management Services): I am surprised that this paper can be cited as a fine example of a professional application of statistics in the public interest. The authors concentrate on a severely flawed data source (numbers killed and seriously injured) which is uneven in definition, subjective, greatly under-reports injuries to pedestrians and cyclists, and is an unknown mixture of the tragic and the trivial — like measuring the money in a collecting box by counting the number of coins irrespective of their denomination. This is justified on the grounds of reduced random variation. Given a monthly *KSI* of 5000, and assuming this derives from 2500 accidents following a Poisson distribution, we would expect the random component in the series to have a standard deviation of 2%. Thus the authors' R-squared of 0.76 indicates a large unexplained component caused by one or more omitted variables.

Previous speakers have drawn attention to the absence of intensity of enforcement of drink driving laws, and the change in the law in effect for most of the test period. An even more important omitted variable is admitted by the authors to have caused large residuals, i.e. the weather. Given the conspicuous effect of weather on accidents and the ready availability of the data, to ignore it in favour of applying an untried and complex statistical method to an obviously mis-specified model seems self-indulgent. It may also explain the completely implausible coefficients on the explanatory variables. Can anyone suppose that a 1% rise in the cost of petrol would reduce casualties by .31% given no change in the miles travelled by cars, or that a doubling of miles travelled (at constant petrol prices) would increase casualties by only 8%? Finally I note that using the unambiguous figures of people killed, the authors claim a saving of only about 5% which is presumably not 'statistically significant'.

Dr O. D. Anderson (Nottingham): As the authors note, when working with economic series, a model that is frequently encountered has the form;

$$(1 - B)(1 - B^T)Z_i = (1 - \theta B)(1 - \Theta B^T)A_i \tag{1}$$

an $IMA(1, 1) \times IMA_T(1, 1)$—the so-called Airline model.

To see how this may be interpreted, consider first a random walk process $\{X_i\}$ contaminated by a purely random sequence of (white noise) observation errors $\{Y_i\}$, to give the recorded process $\{Z_i = X_i + Y_i\}$. Then

$$(1 - B)Z_i = (1 - B)X_i + (1 - B)Y_i = MA(0) + MA(1) = (1 - \theta B)A_i.$$

That is, an $I(1)$ measured with error yields an $IMA(1, 1)$. (Sometimes the "error" is merely due to a rounding of the data.)

Similarly, an $I(1) \times I_T(1)$ contaminated with random noise gives, approximately, the model

$$(1 - B - B^T + B^{T+1})Z_i = (1 - \alpha B - \alpha B^T + \alpha^2 B^{T+1})A_i \tag{2}$$

which is a more parsimonious (and sometimes more appropriate) alternative Airline model. However, the way identification is carried out, makes it likely that the associated correlogram will be read as indicating (1) rather than (2).

Now consider the process of interest, $\{X_i\}$, following the very general model, $ARIMA(p, 1, q) \times ARIMA_T(P, 1, Q)$, and again observe this in the presence of added white noise $\{Y_i\}$. Then we get

$$(1 - B)(1 - B^T)Z_i = ARMA(p, q) \times ARMA_T(P, Q) + (1 - B)(1 - B^T)Y_i$$

and it can be deduced that, under certain conditions, the right of this should almost certainly be identified as $(1 - \alpha B - \alpha B^T + \alpha^2 B^{T+1})A_i$; although, again, current practice would likely yield $MA(1) \times MA_T(1)$. Less restrictive conditions allow the $MA(1) \times MA_T(1)$ to still appear plausible.

When θ approaches unity in the $IMA(1, 1)$ model, the behaviour approaches that of white noise—the $MA(1)$ operator, $(1 - \theta B)$, approximately cancelling with the $I(1)$ differencing, $(1 - B)$. An alternative interpretation is that, in the signal plus noise formulation discussed above, the observation error variance is large compared to that of the underlying process innovation shocks.

These ideas generalise to $ARUMA$ models—ones whose homogeneous nonstationary part, $(1 - B)^d$, is replaced by $U_d(B)$, a polynomial of degree d with all its zeros lying on the unit circle, but not necessarily having all (or any) taking the value unity. Then an $ARUMA$ signal plus random noise will yield a generalised alternative Airline model; an UMA model with an MA part that is just a weak reflection of the $U_d(B)$ operator. Thus $ARUMA(p, d, q)$ + noise $\rightarrow UMA(d, d)$. For instance, if the nonstationary $U_d(B) = (1 + B)(1 + 2B \cos \omega + B^2)$ say, then we would get approximately

$$(1 + B)(1 + 2B \cos \omega + B^2)Z_i = (1 + \alpha B)(1 + 2\alpha B \cos \omega + \alpha^2 B^2)A_i.$$

This explanation for the observed ubiquity of the Airline model was first suggested in Anderson (1980), and is the subject of a report in preparation.

Mr Kuldeep Kumar (University of Kent at Canterbury): The present paper by Professors Harvey and Durbin gives an extremely interesting application of intervention analysis based on structural time series modelling in practice and it also intensifies the technical debate on the use of structural time series modelling and $ARIMA$ modelling in these type of situations.

According to the authors one of the main disadvantages of $ARIMA$ modelling is that its identification procedure is hard to operate and is based on correlogram. Personally I do not agree with the authors to some extent and wish to contribute that during the last decade significant development has taken place in the field of identification of $ARIMA$ and other related models. It is no more true that $ARMA$ models can be identified only by looking at the autocorrelation function and partial autocorrelation function but identification can be done in a more correct way using these latest methods. Some of these methods are based on the generalisation of $PACF$ and ACF. Another important tool for identification is Akaike's

Information Criterion. Seasonal model in this case can be identified by using a more sophisticated technique of extended sample autocorrelation function introduced by Tsay and Tiao (1983).

Can I ask the authors why they have taken the log of the data and why not any other transformation? Authors have stressed that structural models do better than *ARIMA* models. Another alternative possibility is to look at switching *ARIMA* models (one before January 1983 and another after this period) what, if any are the drawbacks of such an approach?

Dr A. Maravall (Bank of Spain, Madrid): I thank the authors for an enjoyable paper and, following their notation, will comment briefly on the use of the BSM as a preferable alternative to *ARIMA* models for the purpose of their paper.

The *BSM* can be seen as the sum of a trend, a seasonal and a white-noise irregular component where the trend is an *IMA*(2, 1) model and the seasonal component follows a model whereby $S(B)\gamma_t$ equals an *MA*(11). Within the *ARIMA* family, the Airline model, for example, can be expressed as the sum of components that follow the same type of models, the difference being that the *BSM* depends on four parameters while the Airline model depends on three. There may be instances where the additional parameter is of help. In their example, however, since $\sigma_\zeta^2 = \sigma_\omega^2 = 0$, the *BSM* simplifies into the sum of a constant, a random walk, a set of seasonal dummy variables and a white-noise component.

The authors suggest that identification of the appropriate *ARIMA* model would have been difficult. I disagree. Since $\nabla_{12} y_t$ has a nonzero mean, the use of a more flexible local trend recommended by the authors is equivalent to taking an additional difference. This leads to the *ACF* of Figure 2, which points to the Airline model. If, instead the *ACF* of $\nabla_{12} y_t$ is considered (Figure 3), starting from a low value for ρ_1, the autocorrelations converge very slowly and are all (except ρ_{12}) positive. This is a clear indication of an *IMA*(1, 1) structure (see Box and Jenkins, 1970, A.6.1), with an additional $MA_{12}(1)$ accounting for ρ_{12}. Thus Figure 3 also suggests an Airline model.

Estimation of the model yields $\Theta = -1$ and hence removing ∇_{12} from the *AR* and *MA* parts of the model, it can then be expressed as the sum of a constant, a random walk, seasonal dummies and a white-noise component, exactly the same decomposition as the one obtained by the authors from the *BSM*. Since the models are the same, why does estimation of the structure rather than the reduced form yield a better estimate of an intervention parameter?

More generally, I sympathize with the idea of a basic or "default" model. But when the basic model is not compatible with the observed *ACF* it is unclear how to extend the *BSM*. On the contrary, from the appropriate *ARIMA* model the full structure can be derived (see, for example, Burman, 1980 or Hillmer and Tiao, 1982). Also, the extension of the *BSM* to the multivariate case is a complicated, still unaddressed, issue. Yet for the set of series considered in the paper, a multivariate *ARIMA* model would have provided an alternative approach worth considering.

Finally, the authors had data on seat belt wearing rates and some other variables. Only the latter are included as explanatory variables although the mandatory seat belt wearing law would have an effect on casualties mostly as a result of increasing the use of seat belts. Hence a natural approach would have been to use the rate of seat belt wearing as an explanatory variable for road casualties and to test for whether the introduction of the law had an effect on that rate.

Dr Emanuel Parzen (Texas A & M University): Professors Harvey and Durbin deserve our admiration and thanks for an excellent and stimulating paper. The numbered points that I would like to add to the technical debate reflect the point of view of my approach to time series modelling (exemplified in Parzen, 1983; Maidment and Parzen, 1984). (I) Why do the authors write equations (2.8) in preference to the Kalman notation for a state space Kalman filter model (chosen in honour of the mentor of system engineers Dr F. G. H. Linear): $Y(t) = H\alpha(t) + e(t)$, $\alpha(t+1) = F\alpha(t) + G\eta(t+1)$. (II) I agree with the author's proposal that state space models are suitable for intervention analysis because they include regression models; therefore intervention could be modeled by $Y(t) = H\alpha(t) + \lambda W(t) + e(t)$ where $W(t)$ is the intervention series. (III) I disagree with the authors that the structural model should be automatically introduced in the initial model identification stage. Structural models must compete not only with Box-Jenkins style *ARIMA* models, but also with *AR*(13) and subset *ARMA*(13, 13) models. Note that after identification *ARMA* models can be expressed as a state space model for the intervention stage of the investigation. (IV) I repeat my proposal that the first step in time series modeling is to determine if the time series is long memory or short memory (for which no transformation such as differencing and/or subtraction of seasonal (monthly) means is required in order to transform a long memory time series to one with short memory). *When a time series is short memory its model form can be* *identified semi-automatically rather than guessed to be a structural model.* (V) That the time series of car drivers killed or seriously injured January 1969–December 1981 (of length $T = 156$) is short memory is indicated by the value of $1 - R^2$; it is 0.24, which indicates short memory (since $0.24 > 8/T = 0.05$). For the model "first differences minus the seasonal means of first differences is a $MA(1)$ with $r(1) = -0.475$ and therefore normalized residual variance equal to 0.64", the diagnostic statistic analogous to $1 - R^2$ has a similar value; it is approximately 0.21, computed from numbers in the paper by 0.64 times $(1 - R^2)/(1 - R_s)$. (VI) Even if one fits a structural model, rather than an *ARMA* model, goodness of fit should be judged in the spectral domain by how well the estimated spectral density $\hat{f}(\omega)$ of $Y(t)$ implied by the model fits the raw sample spectral density $\tilde{f}(\omega)$. I would like to ask the authors if one could approximate an estimated spectral density $\hat{f}(\omega)$ by computing the coefficients of the infinite moving average representation of $Y(t)$ in terms of the standardized residuals $\tilde{v}_t$. The ratio $\tilde{f}(\omega)/\hat{f}(\omega)$ is tested for flatness or whiteness by forming its spectral distribution function or its autoregressive spectral estimator.

Dr D. A. Preece (East Malling Research Station): Section 7.1 mentions puzzling results for back-seat passengers, pedestrians and cyclists. Might not *all* these results become clearer if some account were taken of the *age* of casualties? I'm thinking here particularly of children in the back of a car (who were the subject of a recent BBC television campaign by Miss Esther Rantzen) and of youngsters caught up by the now waning BMX craze. Perhaps sales figures for conventional and BMX-style bicycles might be helpful too? Also, it would be pleasing to be able, some day, to detect benefit from the increased use of reflecting sashes and arm-bands by cyclists and young pedestrians.

Dr W. H. Rutherford (Royal Victoria Hospital, Belfast): The science of traffic accident and injury analysis is a multidisciplinary field. One difficulty with such fields is that it is not easy for those from one discipline to understand the intricacies and technicalities of those from other disciplines. Yet the full understanding must involve each having some appreciation of the other.

I realise that at a meeting of the Royal Statistical Society it is right and proper that much basic understanding of statistical theory and practice should be assumed, and the debate should be at the growing edge of the subject. This appears to me what Professor Harvey and Professor Durbin have done, both using the most advanced statistical analysis to weigh up evidence concerning traffic casualties following seat belt legislation, and also using the assessment of the legislation to be an evaluation of the statistical method.

While I can vaguely appreciate the importance of their whole paper, I have to admit my inability to follow through a single paragraph of the equations and calculations. In this statistical meeting that is of no great import. It seems to me, however, that the engineers, surgeons, psychiatrists, economists and statisticians involved in traffic and injury analysis do also need occasions to meet and exchange ideas when there is time for question and answer at the most simple and basic level, and where we make the effort to be intelligible and to understand across the barriers of our various specialised disciplines. As far as the medical work in this field goes, much of it has, I believe been statistically unsound, and I welcome the intervention of the best available statistic evaluation, and still hope that given time I may come to understand it.

The **authors** replied later, in writing, as follows.

We are grateful to the discussants for their interesting and varied comments. We shall reply point by point.

Dr Chatfield notes the similarity between the basic structural model (*BSM*) and Holt-Winters (HW). We agree that HW has always tended to do quite well in forecasting competitions, and this adds to our confidence that structural modelling will generally give sensible results even though, as Dr Chatfield says, it "... is not yet well-tried". Of course, as Dr Chatfield notes, the great advantage of structural modelling is that prediction MSE's can be calculated, and the model can be extended in a number of ways. As regards model selection, the approach is akin to regression in that explanatory variables and unobserved components are brought into a model on partly *a priori* considerations, and are dropped if they appear to contribute little in the way of explanation. The model selection methodology is best illustrated by reference to some of the applications, such as those in Harvey (1985) and Harvey *et al* (1986). It is noteworthy that in both these cases conventional techniques, i.e. *ARIMA* modelling and regression, gave misleading results. As regards the question of whether or not to take logs, (also raised by Mr Kumar), a check on the chosen specification is provided by our heteroscedasticity and normality diagnostic test statistics. It is worth noting that we experimented with the square root transformation for some series with smaller numbers of casualties; some of the results are given in Durbin and Harvey (1975).

Dr Chatfield also argues that simpler techniques should be used initially in order to get a "feel" of the data. We agree, and in our report to the Department of Transport, graphs of the series and simple before and after comparisons are presented. We would, however, stress that simple comparisons can be misleading and they should not be regarded as a substitute for a more thorough analysis. We shall return to this point when we deal with Professor Ehrenberg's contribution.

Finally Dr Chatfield offers an explanation for the rise in rear seat passengers killed. We are aware of this explanation (see also Mr Jones's comments), but what puzzled us is why a similar rise was not found in the figures for those seriously injured.

We found Dr Tunnicliffe-Wilson's remarks very interesting, particularly his analysis of the pedestrians killed series. However, we still prefer the use of the *CUSUM* in this context, as in our Figure 5, since this is not tied to the notion that there may be step changes in the series. The only point on which we take issue with Dr Tunnicliffe-Wilson is when he says that "*ARIMA* models acknowledge the fact that series arise from complex systems, and seek empirically to capture the dynamics, with no pretence that they can represent the true structure". We wish to stress that a structural model is not attempting to represent the "true structure" as such, but sets out to provide a description of the salient features of the series, to an acceptable degree of approximation. At the same time it imposes constraints which ensure reasonably sensible forecasts.

Professor Priestley quotes us as saying that we are "... not happy with the assumption that differenced series can be assumed stationary ...", and then goes on to point out that the *BSM* is stationary after differencing. This last observation is, of course, true, but the point we were trying to make when we said that we are "... sceptical about the *emphasis* on stationarity of the differenced series", (italics added) is that we are not happy with the idea that the differenced series may be regarded as having been generated by any model within the class of linear stationary processes. We would stress that there is no reason why, in the real world, differenced series should be stationary, and making this assumption could lead to an inappropriate model being chosen from what is a very large class. It is perhaps relevant to refer to Mr Kumar's contribution at this stage since he makes the point that there are several techniques, apart from the correlogram, which can be used to identify *ARIMA* models. We are aware of these techniques but would argue that they may merely enable one to choose an inappropriate *ARIMA* model in a more sophisticated way.

Professor Priestley also suggests that we should have experimented with different change points and dynamics for the intervention variable. In sub-sections 5.2 and 5.5 we discussed in some detail why this should not be done in an unstructured way, and we presented some ideas on what we felt was a suitable approach from the methodological point of view.

Professor Burman asks about the relative fits of the *BSM* and airline models. In general, the models are very close for this kind of data and in certain special cases, for example car drivers *KSI*, they are actually equivalent; see also Maravall (1985). As regards the *ML* estimation of the *BSM*, a full description can be found in the paper by Harvey and Peters (1984).

Mr Stern, Mr West-Oram, Mr Stewart and Dr Adams referred to the effects of alcohol consumption on road accidents and there is no doubt that this is an important factor. However, the interpretation of the relevant statistics is a complex matter, and there are no readily available monthly data on alcohol consumption of car drivers. The evidence produced by Dr Adams is somewhat indirect and in fact if one looks at the series to which he refers, 1982 appears to be an unusually high year compared to 1980 and 1981. The change in the law on evidential breath testing took place in May 1983 and it is difficult to discern a change in the casualty figures at that time although there may have been an anticipatory effect. The study of the effects of alcohol is clearly an important area for future research.

As regards Dr Adams's remarks about fatalities versus *KSI* data, we stick to our belief that the *KSI* analyses provide valuable information and refer the reader to Mr Dale's comments and to the additional data he provided in his contribution to the discussion. Even on the basis of our analysis of those killed we cannot accept Dr Adams's suggestion that the balance of the evidence "tilts strongly in favour of the conclusion that there has been no net life-saving benefit attributable to the seat belt law".

Mr Scott asks whether the stochastc trend component will "mop > ›" variation due to factors which have been excluded from the model. The answer is yes, though how successful the mopping up operation is will depend on the nature of the changes in these excluded factors. Of course if the excluded factors can be measured, it is better to try to bring them into the model. It may be that there is a better way of accounting for the changes which took place around 1974 than the petrol price variable which we used.

However, we felt this was a reasonably parsimonious way of taking account of changes which occurred both then and in 1979. As regards the point about whether a simpler model fitted to a shorter time series might be adequate, the answer is, in general, no. It would, for example, be inappropriate to use a regression model with time as the explanatory variable if an analysis of the data had shown it to be generated by a random walk.

It is difficult to respond to Professor Ehrenberg's tirade since it contains little in the way of relevant rational argument. Furthermore the points he does make seem naive. For example, he claims that a simple before and after comparison gives virtually the same results as our analysis. This is true for the car drivers *KSI* but it is not true for certain other series such as pedestrians killed. The fact that a simple before and after comparison *sometimes* gives the same result as a model which allows for a trend and explanatory variables does not imply that it *always* gives the same result. At the simplest level it is easy to show that if the data contain a trend, a before and after comparison can give an intervention effect with the wrong sign!

We welcome Professor Smith's suggestions about the closer examination of the information in the data about λ, notably the plotting of the profile likelihood and the posterior distribution, and hope someome will follow up these suggestions using our data set.

We found Professor Cane's remarks on practical aspects of road accident data most interesting. In particular, the point about the buffer effect of front seat on rear seat occupants could be relevant and was new to us. We thought about including a speed variable at the outset of the investigation, but since the available data were macro-data which showed a slow steady increase throughout the period of interest we realised that it would not be possible to disentangle the effect of speed from that of a general trend in the series.

Mr Stewart criticizes the goodness of fit of our model on the basis of the assumption that the data follow a Poisson distribution. Perhaps this is an indication that the Poisson assumption is an inappropriate one for complex events such as road casualties! If Mr Stewart can find a model with a standard deviation of 2% we'd be very interested to see it. We agree that it would be interesting to examine the effect of introducing weather variables as explanatory variables. Because of the limitations of time available we were unable to undertake this and we hope others will do so using our data set. A sophisticated treatment will be needed since while the initial effect of a deterioration of weather conditions on the accident rate would be expected to be adverse, when the weather deteriorates beyond a certain point the accident rate goes down, presumably because fewer people venture onto the roads while those that do so recognise the need for greater care than usual. Furthermore, variations in accidents are presumably caused not just by precipitation and temperature, but by some combination of both. A given level of precipitation and a given average temperature in a month may or may not lead to hazardous road conditions depending on the timing.

Dr Maravall raises the question of generalizing the basic structural model to the multivariate case. The theory underlying this generalization is set out in Harvey (1985b, 1985c) and an algorithm for estimating multivariate structural models in the frequency domain has been developed by one of our students, Javier Fernandez-Macho; see Fernandez-Macho (1986). The software for handling explanatory and intervention variables is currently being written. We agree with Dr Maravall that a multivariate approach is well worth considering for problems like the seat belt one, and we are optimistic about the potential value of employing multivariate structural models.

We agree with Professor Parzen that structural modelling has to compete not only with *ARIMA* modelling but also with other types of modelling such as subset *ARIMA* (13, 13) modelling. As regards tests in the frequency domain, we agree that the spectral distribution function, as estimated by the cumulated periodogram, is a useful diagnostic tool. However, it cannot be used to provide a formal goodness of fit test since Durbin (1975) has shown that for dynamic models it has an intractable parameter-dependent limiting distribution.

In reply to comments by Dr Rutherford and Mr Stern on the readability of the paper to laymen, we wish to point out that the paper is addressed primarily to statisticians. We refer anyone requiring a simpler exposition to our report to the Department (Durbin & Harvey, 1985), in which we went to considerable pains to try to make our analyses and results accessible to non-specialists. There is a copy of our report in the RSS Library.

REFERENCES IN THE DISCUSSION

Akaike, H. (1974a) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Statist. Math.*, **26**, 363–387.

——(1974b) Stochastic theory of minimal realisations. *IEEE Trans. Autom. Control, AC-19*, 667–673.
Anderson, O. D. (1980) *Forecasting Public Utilities*. Amsterdam: North-Holland.
Barnard, N. R., Barwise, T. P. and Ehrenberg, A. S. C. (1986) *Re-interviewing in Attitude research: Early Results*. The Annual Conference. London: Market Research Society.
Burman, J. P. (1980) Seasonal adjustment by signal extraction. *J. R. Statist. Soc.* A, **143**, 321–327.
Chatfield, C. (1978) The Holt—Winters forecasting procedure. *appl. Statist.*, **27**, 264–279.
——(1985) The initial examination of data (with Discussion). *J. R. Statist. Soc.* A, **148**, 214–253.
Durbin, J. (1975) Tests of model specification based on residuals. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) pp. 129–143. Rotterdam: North-Holland.
Ehrenberg, A. S. C. (1986) *Advertisers or Viewers Paying?* London: Admap Publications (Admap Monograph).
Ehrenberg, A. S. C. and Wakshlag, J. (1986) Repear viewing in the Boston Panel. In Proceedings of ARF Workshop. New York: Advertising Research Foundation.
Ehrenberg, A. S. C. and England, L. R. (1986) The design of the 1986 LBS pricing experiment. London Business School: CMaC Working Paper.
Fernandez-Macho, J. (1986) On the estimation and testing of the multivariate basic structural model. Unpublished paper, London School of Economics.
Goodhardt, G. J., Ehrenberg, A. S. C. and Chatfield, C. (1984) The Dirichlet: a comprehensive model of buying behaviour (with Discussion). *J. R. Statist. Soc.* A, **147**, 621–655.
Hansard (1986) House of Lords, Vol. **470**, No. 29, col. 69.
Harvey, A. C. (1985c) Analysis and generalisation of a multivariate exponential smoothing model. *Manag. Sci.*, **32**, 374–380.
Harvey, A. C., Henry, S. G. B., Peters, S. and Wren-Lewis, S. (1986) Stochastic trends in dynamic regression models: an application to the employment—output equation. *Econ. J.*, to appear.
Hillmer, S. C. and Tiao, G. C. (1982) An ARIMA-model based approach to seasonal adjustment. *J. Amer. Statist. Ass.*, **77**, 63–70.
Lievesley, D. and Waterton, J. (1986) Advantages and limitations of a panel approach in an attitude survey. Proceedings of the Annual Conference. London: Market Research Soc.
Lowne, R. *et al.* (1984) The effect of the UK seat belt legislation on restraint usage by children. In *Proc. SAE Congress*, pp. 369–380. Detroit, Michigan 840526.
Maidment, D. R. and Parzen, E. (1984) Casade model of monthly municipal water use. *Water Resources Res.*, **20**(1), 15–23.
Makridakis, S. *et al.* (1982) The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. of Forecasting*, **1**, 111–153.
Parzen, E. (1983) Time series model identification by estimating information, memory and quantiles. *Questo*, **7**, 531–562.
Tsay, R. S. and Tiao, G. C. (1983) Identification of multiplicative ARMA models for seasonal time series. Technical Report No. 7, Statistics Research Center, Graduate School of Business, University of Chicago.
Voltaire (1750) *Histoire de Charles XII, Roi de Suède* (no date or place of publication).

As a result of the ballot held during the meeting of 5th March, the following were elected Fellows of the Society.

Chudy, Bernard
Davies, Richard B.
De Morgan, Richard M.
Else, Elisabeth R.
Eltinge, John L.
Gilmour, Tom
Harrison, M. R.
Laing, John B. G.
Lee, Patrick J.
Legge, Kathryn
Mead, Andrew
Mohamad, Hadi A.
Saidian, Masouda
Struthers, Lesley P. L.
van der Ploeg, Carol E.
Williams, Linda A.