Deliverable D7.7: "Multivariate time series analysis of SafetyNet data"

Contract No: TREN-04-FP6TR-S12.395465/506723
Acronym: SafetyNet
Title: Building the European Road Safety Observatory

Integrated Project, Thematic Priority 6.2 "Sustainable Surface Transport"

Project Co-ordinator:
Professor Pete Thomas
Vehicle Safety Research Centre
Ergonomics and Safety Research Institute
Loughborough University
Holywell Building
Loughborough, LE11 3UZ

Organisation name of lead contractor for this deliverable:
SWOV Institute for Road Safety Research

Due Date of Deliverable: 30/05/2007

Submission Date: 13/06/2007

Report Author(s): J.J.F. Commandeur (SWOV), F.D. Bijleveld (SWOV), R. Bergel (INRETS)

Project Start Date: 1st May 2004 **Duration:** 4 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002 -2006)		
Dissemination Level		
CO	Public	

Table of Contents

Preface

This deliverable provides an application of theories and methods documented in Deliverables 7.4 and 7.5 of work package 7 of the SafetyNet project. In this deliverable, use of select analysis techniques is demonstrated through real world road safety analysis problems, using aggregate data which may not be available yet in SafetyNet databases. The prime goal of the analysis in this deliverable however is to demonstrate the analysis techniques, their features and their suitability to answer road safety questions. Consequent on this, the data used in the analysis was selected primarily for their demonstrative usefulness.

This deliverable demonstrates the use of time series analysis techniques. In particular, structural time series models are developed and demonstrated for France and the Netherlands, as well as disaggregated models for two types of networks in France, and disaggregated models for several accident types in the Netherlands. It is demonstrated how road safety developments of the traffic volume, the number of accidents and the number of fatalities can be linked to the developments of exposure, accident risk and accident severity, estimated through their unobserved components: their trend (level and slope) and their seasonals. Some interpretations are given. In addition, the performance of the time series model is compared to the performance of one classical alternative: the vectorial regression model.

Chapter 1 - Introduction

In this deliverable we present the results of the multivariate time series analysis of three levels of the road risk process, as defined in the road safety field:

- vehicle kilometres (or a proxy for vehicle kilometres), from which the unobserved exposure to risk (the real and unknown number of vehicle kilometres travelled[1] is derived,
- the number of injury accidents, from which the unobserved accident risk (the number of injury accidents per vehicle kilometre) is derived,
- the number of fatalities, from which the unobserved fatality risk (the number of fatalities per injury accident) is derived.

We investigate the following relations between these three levels of the road risk process :

number of fatalities =
exposure x (injury accidents/exposure) x (number of fatalities/injury accidents) =
exposure x accident-risk x fatality-risk,

referring to Appendix 1 for the technical details of how these relations are handled using a multivariate local linear trend with seasonal model. We refer to Section 3.6 of the Methodology report D7.4 of the SafetyNet project for a discussion of the univariate local linear trend with seasonal model, while the multivariate approach presented here is discussed more extensively in Bijleveld, Commandeur, Gould and Koopman (2005, forthcoming).

There is a similarity between the models presented here and the well-known DRAG family of models as discussed in Gaudry (1984), and Gaudry and Lassarre (2000), since they also model these three levels of the road risk process: exposure (Demande Routière), accident risk (Accidents) and accident severity (Gravité). The main differences in the models discussed here, compared to the DRAG models, are that:

- the structural time series framework (Harvey, 1989; Durbin and Koopman, 2001) is used instead of the autoregressive framework (Box and Jenkins, 1976) in DRAG models,
- the model can be both descriptive and explanatory, while DRAG models are meant to be explanatory,
- the three levels of the road risk process are analysed simultaneously instead of in three separate steps in the existing DRAG models.

In this deliverable, we shall focus on the descriptive property of the model, and not on the explanatory one. However, as in the DRAG approach, in this approach explanatory variables can be added at all levels of the road risk

[1] Other measures of exposure to risk could have been chosen in the place of the number of vehicle kilometres travelled on a definite network (see Yannis et al, 2005). This one is the most commonly used, due to availability reasons.

process, thus making it possible to investigate both the direct and the indirect effects - via the exposure to risk - of such variables on the developments of the two types of road risk: the accident risk and the fatality risk.

This structural modelling applied to road safety datasets addresses latent - or, in other words, hidden or unobserved - variables. More specifically, each of the three latent variables exposure, accident-risk and fatality-risk is modelled as a sum of its unobserved components: its trend (level and slope), its seasonal component and its residual disturbance term. In that sense, this modelling is qualified as structural modelling.

Structural time series modelling has already been used by researchers in road safety who applied it at a European level (Cost 329, 2004). Lassare (2001) analysed and compared aggregate road safety trends in ten European countries. However, although some attempts of multivariate analysis were made in Bergel (1997, 2004) and in Cost329 (2004), most of the approaches remained univariate, and aimed at analysing the developments of aggregate numbers of accidents and aggregate numbers of fatalities, in relationship to a risk exposure indicator. No global approach of the road risk process, handling its three dimensions at the same time, was conducted nor applied on European road safety data yet, with the help of structural models.

An important aspect of the multivariate structural time series model is that the covariance structure of the disturbances allows the various unobserved components (levels, slopes, and seasonals) to be correlated, which is not the case when univariate time series models are used independently of one another. This is a natural thing to do because developments in exposure to risk, accident risk and fatality risk typically are subject to the same overall environment, which is the traffic process, and so a multivariate model will seek to link them together. But the multivariate approach also opens up the possibility of investigating whether developments in risk have common unobserved components.

For deliverable D7.7 of the SafetyNet project, we apply this multivariate structural time series model both at the national level of European countries (in order to compare the road safety developments in these countries), and on disaggregated data within a country. In all cases quarterly data are used.

Specifically, the multivariate time series model is used to analyse national data from France and from the Netherlands. Results of these two analyses are presented in Sections 2 and 3, respectively. We also apply the three-level model to two different types of French roads, and to several types of Dutch severe injury accidents. The results of these analyses are presented in Sections 4 and 5.

For the analyses presented in this deliverable we have linked exposure data, accident data, and victim data. We expect that in due time such data will also become available in SafetyNet, certainly at a national level of the European countries for which data are being and will be collected as a result of the efforts of WP1 and WP2 of SafetyNet. Moreover, if data should not yet be available it is sometimes still possible to use proxies, as will be shown in the analysis performed of the dataset retained for the whole of France and discussed in Section 2.

The relevance of the models presented in this deliverable for policy makers is that they are well suited for handling the uncertainties and noise in time series data collected in road safety, as well as for handling the time dependencies between such observations. They allow to provide optimal answers to questions of policy makers such as:

- what are the developments in road safety while disentangling exposure, accident risk, and fatality risk?
- what is the effect of a measure on the developments in road safety?
- how can we expect road safety to develop in the future?
- should current or past road safety figures be considered to represent a break from the past or not?
- how do the risks of different parts of the traffic system compare to each other as they evolve over time?

All analyses discussed in this deliverable were performed in Ox (Doornik, 2001) and SsfPack (Koopman, Shephard and Doornik, 1999).

Chapter 2 - National road safety developments in France

For France, we have simultaneously analysed quarterly data for the period 1987-2000 concerning:

- the oil sales[2] for the whole of France, used as a proxy for the exposure on the whole French network, as the number of vehicle kilometres driven is only available for main roads and motorways in France,
- the total number of injury accidents,
-. the total number of people being killed in those accidents.

It is important to note that, concerning the dataset used for France, the definition of an injury accident is: an accident in which there is at least one victim - whether killed (after being hospitalised less than 6 days), whether seriously injured (hospitalised more than 6 days) or lightly injured (hospitalised less than 6 days, or even not hospitalised at all).

The analysis of these three series has first been performed with a multivariate local linear trend with seasonal model. One particular restriction of this general model which turns the model into a classical vectorial regression model is also applied to the same dataset. Likelihood ratio tests were conducted which demonstrated the superiority of the unrestricted multivariate model.

2.1. The multivariate stochastic model

The analysis of these three series with a multivariate local linear trend with seasonal model, which requires the estimation of a total of 24 hyperparameters[3] gives the following results.

[2] Proxy variables should only be used as a last resort, which is the case here for the oil sales in place of the number of vehicle kilometres driven.

[3] Four sets of disturbances are allowed to be correlated: observation disturbances (one for oil sales, one for the total number of injury accidents and one for the total number of people being killed in those accidents), disturbances of the level components (one for exposure to risk, one for accident risk and one for fatality risk), disturbances of the slope components (one for exposure to risk, one for accident risk and one for fatality risk) and disturbances of the seasonal components (one for exposure to risk, one for accident risk and one for fatality risk). Each set of disturbances needs 6 hyperparameters.

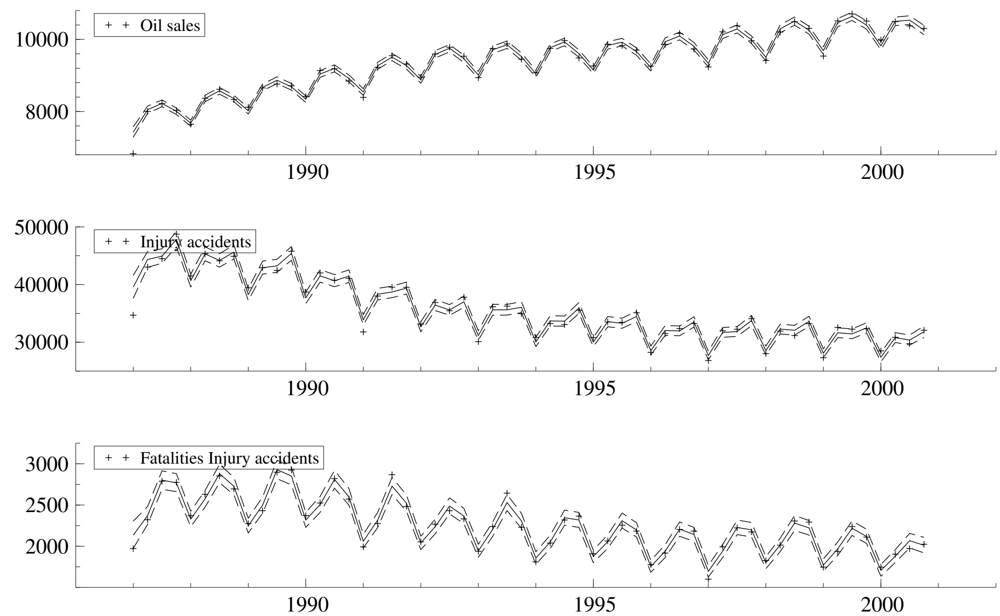

Figure 2.1. Smoothed predictions (uninterupted lines) for quarterly oil sales, injury accidents, and fatalities for the whole of France, including 95% confidence limits (between dashed lines).

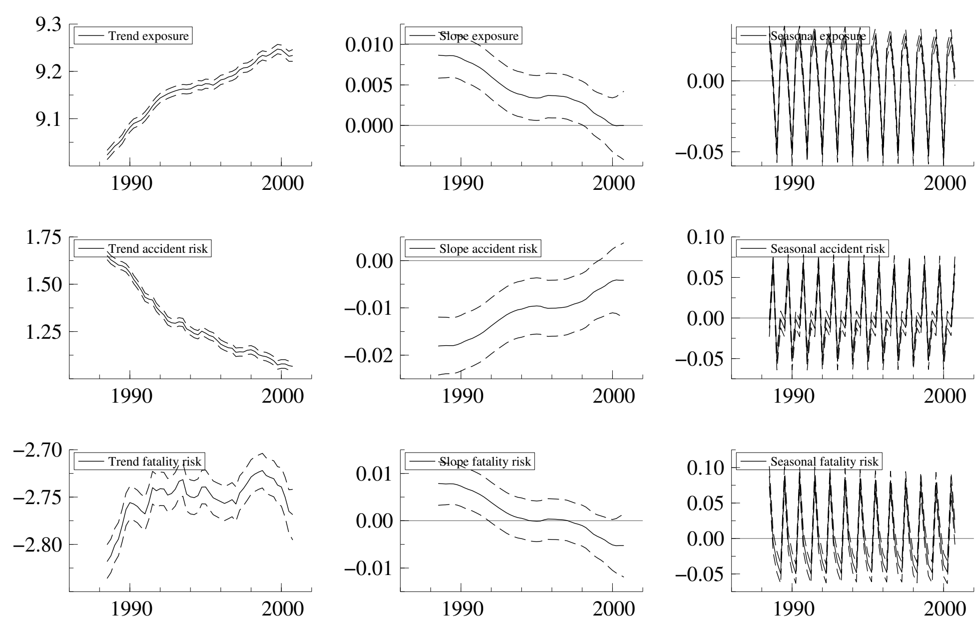

Figure 2.2. Smoothed trend, slope and seasonal components for the whole of France (uninterupted lines), including 95% confidence limits (between dashed lines).

Figure 2.1 shows the development of the three observed variables (oil sales, the number of injury accidents and the number of fatalities), estimated with the multivariate model for the period 1987-2000, using a smoothed prediction[4].

Figure 2.2 shows the main unobserved components (trend - level and slope - and seasonals, in columns) of the three latent variables (the logarithm of the exposure to risk, accident risk and fatality risk, in rows), estimated with the multivariate model for the period 1987-2000. For each latent variable, the estimated value is the sum of the trend value (level and slope) and of the seasonal value. The slope value helps to understand the development of the trend over the period.

As can be seen in the second graph in the first row of Figure 2.2, the time-varying slope component for exposure is non-negative throughout the period 1987-2000, indicating that the exposure in France has continuously increased. Moreover, since the non-negative value of the slope becomes smaller towards the end of the period, the rate of increase in exposure was largest at the beginning of the period, and smallest at the end of the period. Similarly, inspection of the second graph in the second row of Figure 2.2 shows that the slope component for accident risk is negative throughout the time period, but that its absolute value becomes smaller and smaller as time proceeds. At the end of the period the increase may even have stalled. This means that the accident risk in France has continuously decreased, but that the rate of decrease is levelling off towards the end of the period. Finally, the fatality risk has first increased until about 1994, but has started to decrease after 1997, as follows from inspection of the second graph in the bottom row of Figure 2.2. To be accurate, these comments apply to the seasonally corrected[5] exposure, accident risk and fatality risk developments.

Based on the seasonal components in the last column of Figure 2.2, in France the exposure and the fatality risk are always the largest for the third quarter of the year, while the accident risk is always the largest during the last quarter of the year, as is shown more clearly in Figures 2.3a and 2.3b. It may be noted that the seasonal pattern of 1987 (2.3a) is very similar to the average pattern for the whole of the 14 years of the period (2.3b).

[4] Estimations of the observed variable can be provided with two kinds of predictors: the smoothed predictor, which was used here and the filtered predictor. (see Harvey (1989))

[5] Seasonal variation removed

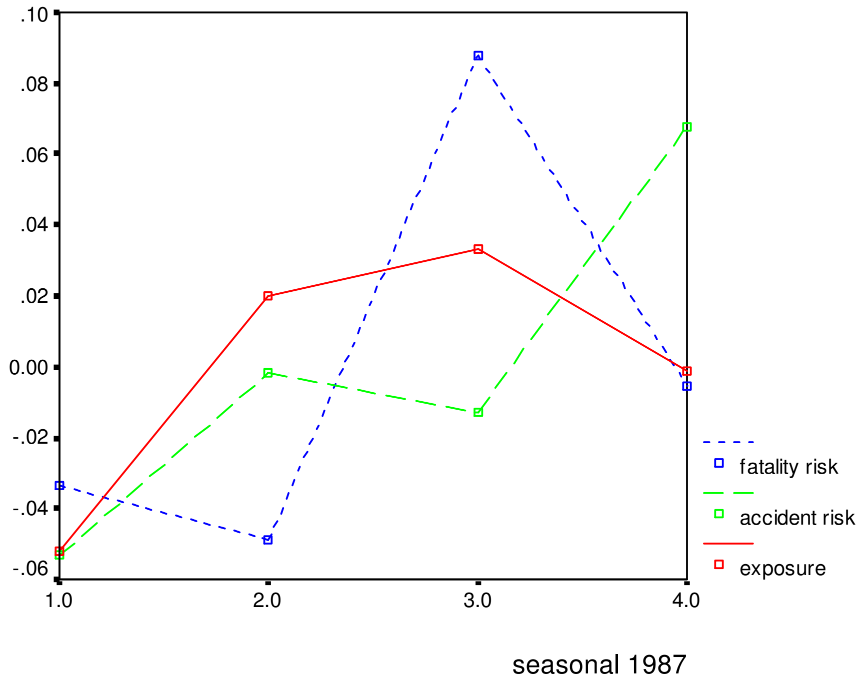

Figure 2.3.a Quarterly seasonals in 1987 for the whole of France.

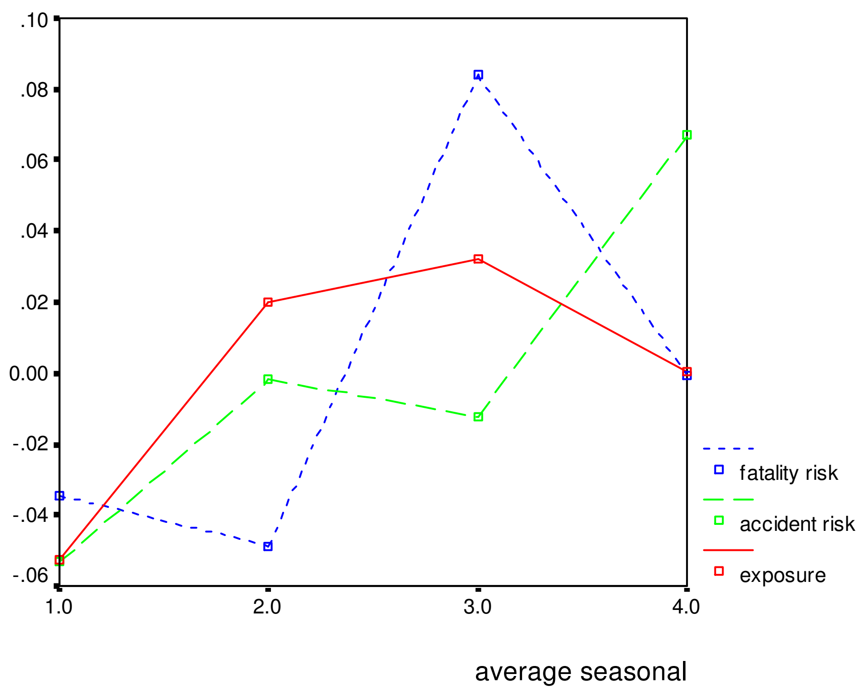

Figure 2.3.b Average quarterly seasonals for the whole of France.

All of the model assumptions concerning independence, homoscedasticity and normality of the standardised prediction errors are satisfied in this analysis. At convergence the value of the log-likelihood function for this stochastic model equals 330.460. The estimated variance matrices of this analysis are given in Appendix 2.

2.2. Comparison with the multivariate deterministic model

Finally, to show the superiority of the multivariate structural time series analysis compared to the use of a multivariate linear regression model we now present some of the results obtained in an analysis with deterministic (fixed) dynamics of these same time series. Since the variance-covariance matrices of the level, slope and seasonal disturbances are all fixed on zero in this case, the analysis only requires the estimation of six hyperparameters (i.e., for the variance-covariance matrix of the observation disturbances).

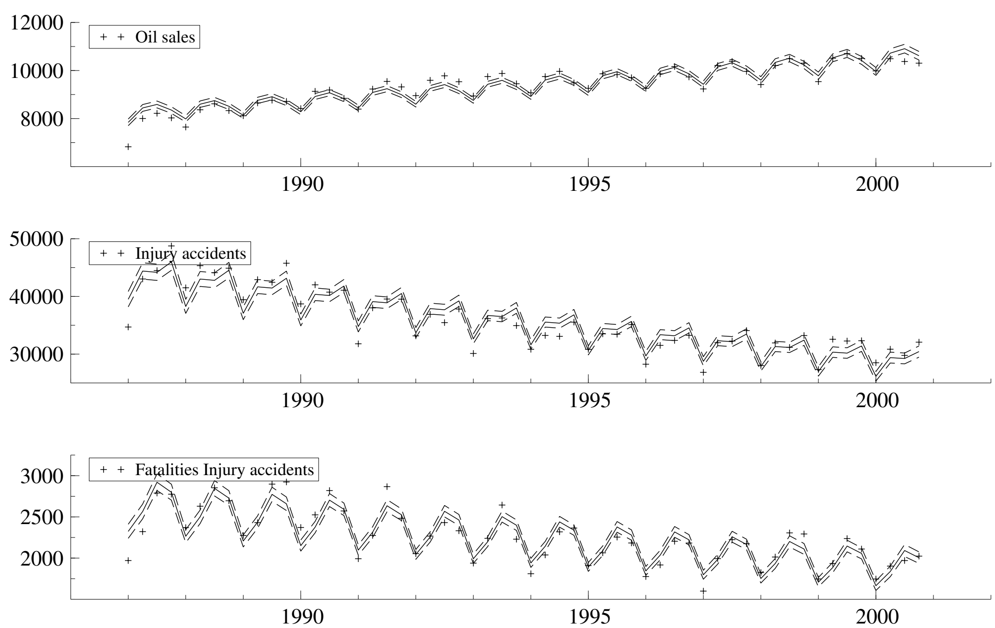

Figure 2.4. Smoothed predictions for oil sales, injury accidents, and fatalities of multivariate linear regression for the whole of France, including 95% confidence limits.

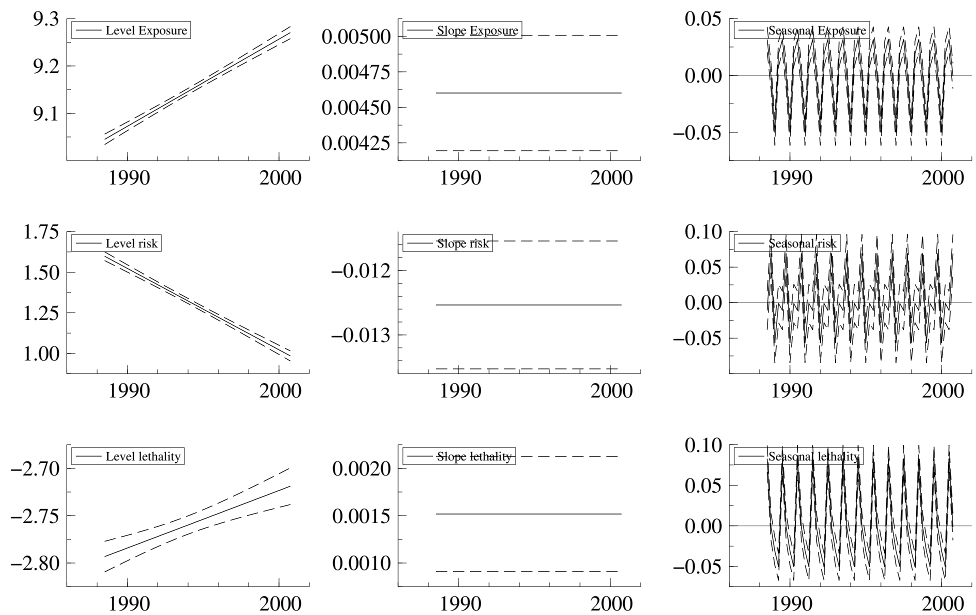

Figure 2.5. Smoothed trend, slope and seasonal components of multivariate linear regression for the whole of France, including 95% confidence limits.

As Figure 2.5 clearly indicates, the trends for exposure, accident risk, and fatality risk are all now simple regression lines, with a fixed slope and intercept, and the seasonal components no longer change over time. A main difference with the previous analysis is that the assumption of independence and the assumption of homoscedasticity are now violated for all three residuals, whether related to oil sales, injury accidents or fatalities. Only the assumption of normality is not rejected for all three residuals. As can be seen in Figure 2.4, more observations fall outside the 95% confidence interval for the smoothed predictions than in Figure 2.1.

At convergence the value of the log-likelihood function for the completely deterministic model equals 270.913. Applying the likelihood ratio test to the values of the likelihood function of the stochastic model and of the deterministic model, we find that

$$2(\log(L_u) - \log(L_r)) = 2(330.460 - 270.913) = 119.094\,.$$

Since this value of the likelihood ratio test is much larger than

$$\mathrm{X}^2_{(0.05;18)} = 28.869\,,$$

where (24 - 6) = 18 is the number of restrictions, the improvement in fit of the stochastic model compared to the deterministic model is significant with $p < 0.05$.

Another way of looking at the differences between the two analyses is by considering the forecasts obtained with the two models, as shown in Figures 2.6 and 2.7.

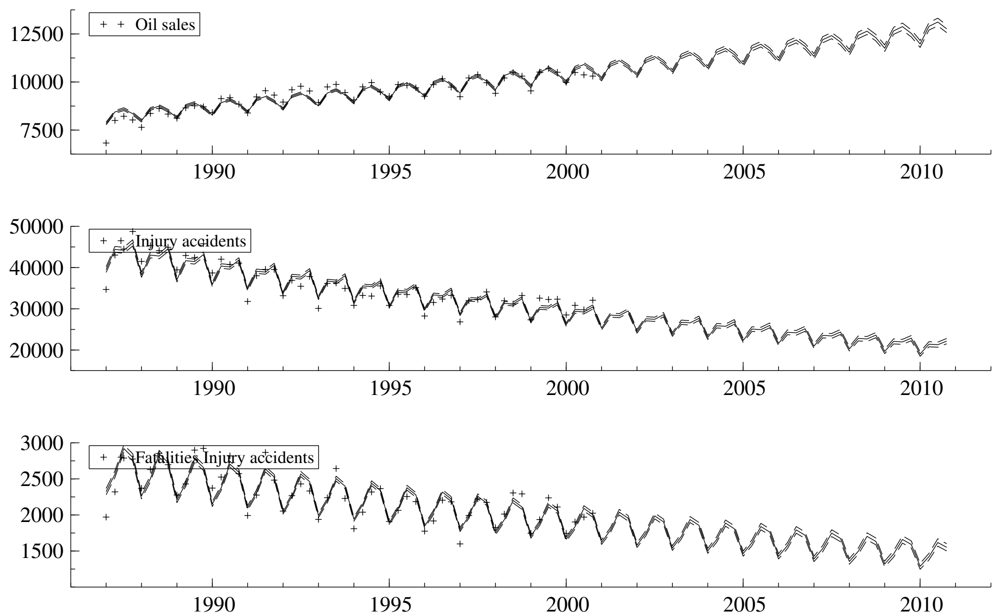

Figure 2.6. Forecasts of deterministic three-level model applied to the whole of France, including 68% confidence limits.

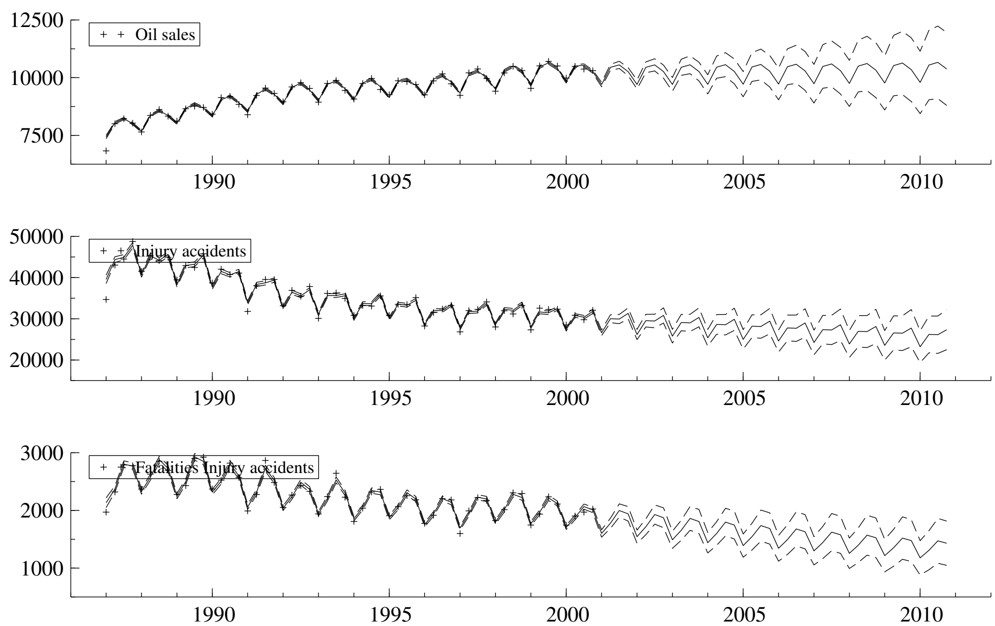

Figure 2.7. Forecasts of stochastic three-level model applied to the whole of France, including 68% confidence limits.

Comparing the forecasts in Figures 2.6 and 2.7 highlights the considerable difference in the values of the confidence limits around these forecasts. These limits are much more conservative in the latter than in the former figure, indicating that forecasts obtained with the multivariate linear regression model result in a false sense of certainty about future developments in road safety.

Chapter 3 - National road safety developments in the Netherlands

For the Netherlands, we have simultaneously analysed quarterly data for the same period as in France (1987-2000) concerning:
- the vehicle kilometres of drivers of all vehicles, thus *including* non-motorized vehicles, for the whole of the Netherlands, based on survey data,
- the total number of injury accidents (resulting in people being killed or seriously injured),
- the total number of people being killed in these accidents.

For the Netherlands, the following definition of an injury accident applies: an accident in which there is at least one victim - whether killed (after being hospitalised less than 30 days), or seriously injured (hospitalised more than one day).

In this chapter, the analysis of these three series has first been performed with a multivariate local linear trend with seasonal model, and then with a multivariate regression model. As in Section 2, a likelihood ratio test is used to demonstrate the superiority of the unrestricted multivariate model.

3.1. The multivariate stochastic model

The multivariate analysis of these three series with a multivariate local linear trend with seasonal model, which requires the estimation of a total of 24 hyperparameters (i.e., six parameters for the variance-covariance matrix of observation disturbances, and 18 parameters for the variance-covariance matrices of the level, slope and seasonal disturbances), gives the following results.

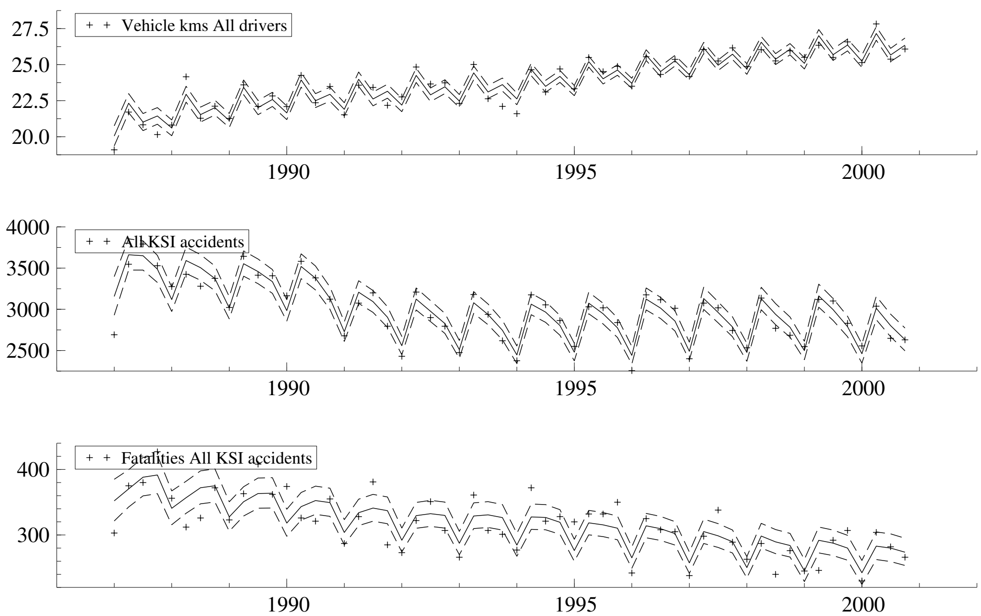

Figure 3.1. Smoothed predictions for quarterly vehicle kilometres, KSI accidents, and fatalities for the whole of the Netherlands, including 95% confidence limits.

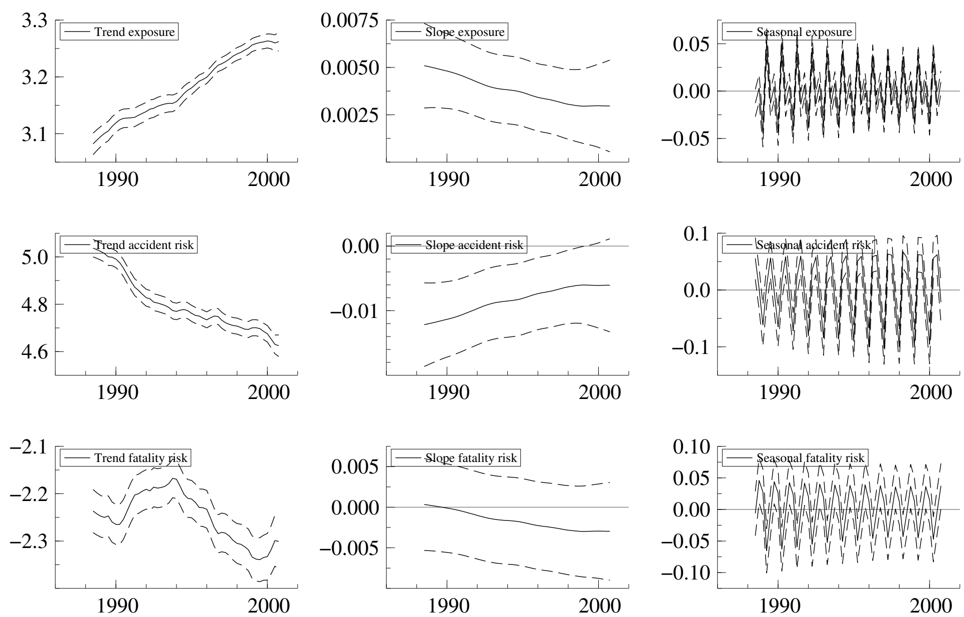

Figure 3.2. Smoothed trend, slope and seasonal components for the whole of the Netherlands, including 95% confidence limits.

Just as in France, the seasonally corrected exposure in the Netherlands has continuously increased, while the seasonally corrected accident risk has continuously decreased, as indicated by the values of the corresponding slope components. In the Netherlands the fatality risk has continuously decreased from 1990 onwards, while in France it first increased until about 1994, only starting to decrease after 1997.

Inspection of the confidence limits of the slope components in Figure 3.2 shows that both the upper and lower bounds are consequently located on the positive side of the horizontal axis for exposure, and that both the upper and lower bounds are (almost) consequently located on the negative side of the horizontal axis for accident risk. However, the upper bound of the confidence interval of each individual slope value for fatality risk is consequently located on the positive side of the horizontal axis while the lower bound is consequently located on the negative side of the horizontal axis. This means that the sign of the development in the slope of the fatality risk in the Netherlands is more uncertain than the signs of the developments of the slopes of the exposure and risk components.

Inspecting the seasonal components in Figure 3.2, for the whole of the Netherlands the exposure is the largest for the second quarter of the year, the accident risk is largest during the third quarter of the year (although the second quarter becomes more and more similar to the third quarter towards the end of the period), while the fatality risk is largest during the last quarter of the year. This can be seen more clearly in Figure 3.3a which shows graphs of the three quarterly seasonals for only one year (i.e., 1987, the first year in the series). It may noted that these differences between quarters are also found when we calculate the *average* seasonals over the 14 years considered in the analysis (Figure 3.3b).

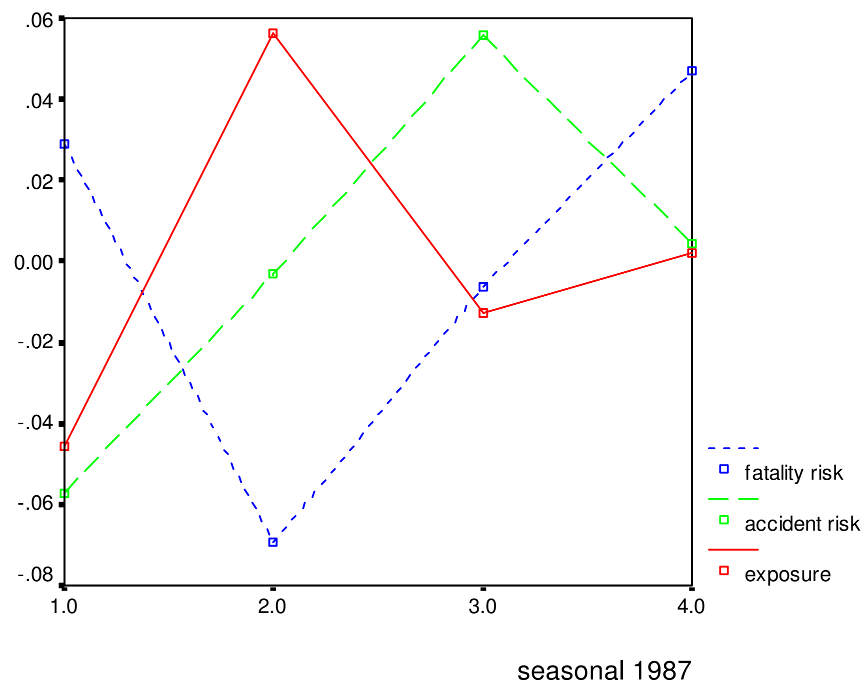

Figure 3.3.a. Quarterly seasonals in 1987 for the whole of the Netherlands.

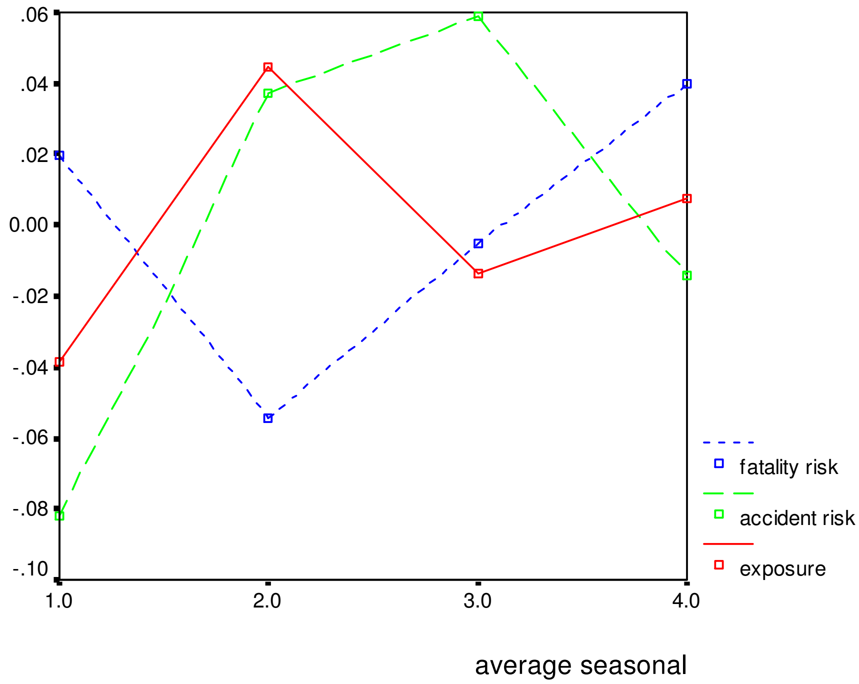

Figure 3.3.b. Average quarterly seasonals for the whole of the Netherlands.

The assumptions concerning independence, homoscedasticity and normality of the standardised prediction errors are all satisfied in this analysis. At convergence the value of the log-likelihood function for this stochastic model equals 219.746. The estimated variance matrices of this analysis are given in Appendix 3.

It is of some interest to note that the eigenvalue decomposition of the variance matrix of the level disturbances shows the third largest (smallest) of the three eigenvalues to be very small. The same applies to the second and third largest eigenvalues of the variance matrix of the slope disturbances, and to the second and third eigenvalues of the variance matrix of the seasonal disturbances. This implies that the three level components are possibly driven by only two sources of error, while the three slope components and the three seasonal components are both possibly driven by only one source of error. This means that the same analysis could be repeated applying the just mentioned rank restrictions on the variance matrices of the state disturbances without affecting the value of the log-likelihood function much.

It is not possible to compare the developments of the three levels for the Netherlands and France in absolute terms. For example, the range on the *y*-axis of the trend of the unobserved fatality risk is between exp(-2.4) = 0.091 and exp(-2.1) = 0.122 for the Netherlands in Figure 3.2, suggesting a range of 0.09 to 0.12 persons killed per injury accident in this country. For France, on the other hand, the range on the *y*-axis of the trend of the unobserved fatality risk in Figure 2.2 is between exp(-2.85) = 0.059 and exp(-2.7) = 0.067, suggesting a range of only 0.06 to 0.07 persons killed per injury accident, and therefore a smaller accident severity than for the Netherlands. These smaller numbers are however at least partly explained by the different definitions used in the two analyses for an injury accident, the definition for France being much more liberal (and therefore involving many more, less severe accidents) than for the Netherlands. However, the use of a correcting factor applied to the number of fatalities as defined in France before 2005 would allow to correct for that difference to some extent .

3.2. Comparison with the multivariate deterministic model

To show the superiority of the above analysis compared to the use of a multivariate linear regression model we now present some of the results obtained in a deterministic analysis of these same time series. Since the variance-covariance matrices of the level, slope and seasonal disturbances are all fixed on zero in this case, the analysis only requires the estimation of six hyperparameters (i.e., for the variance-covariance matrix of the observation disturbances).

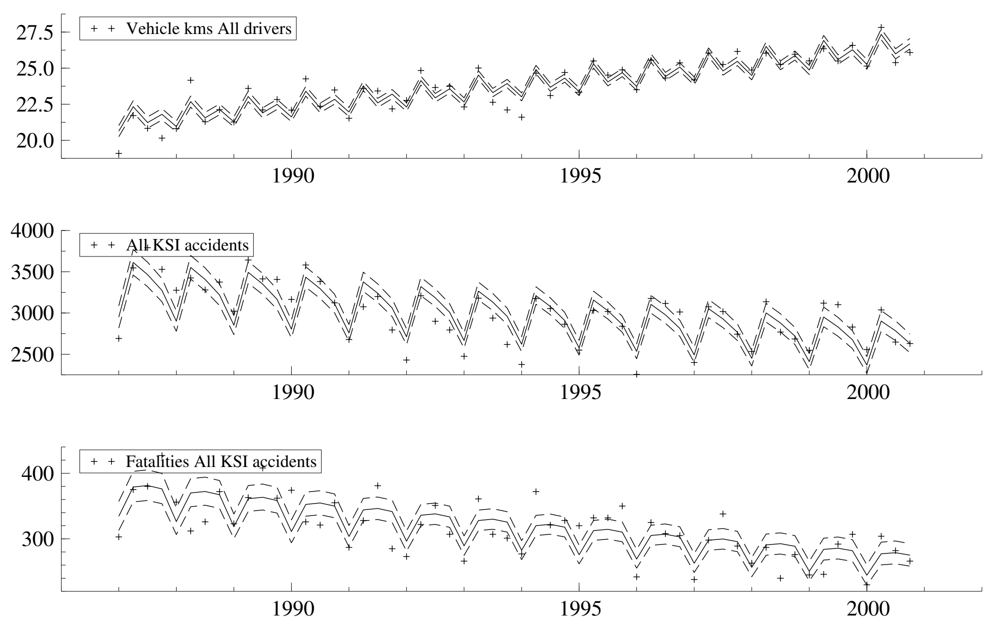

Figure 3.4. Smoothed predictions for quarterly vehicle kilometres, KSI accidents, and fatalities of multivariate linear regression for the whole of the Netherlands, including 95% confidence limits.

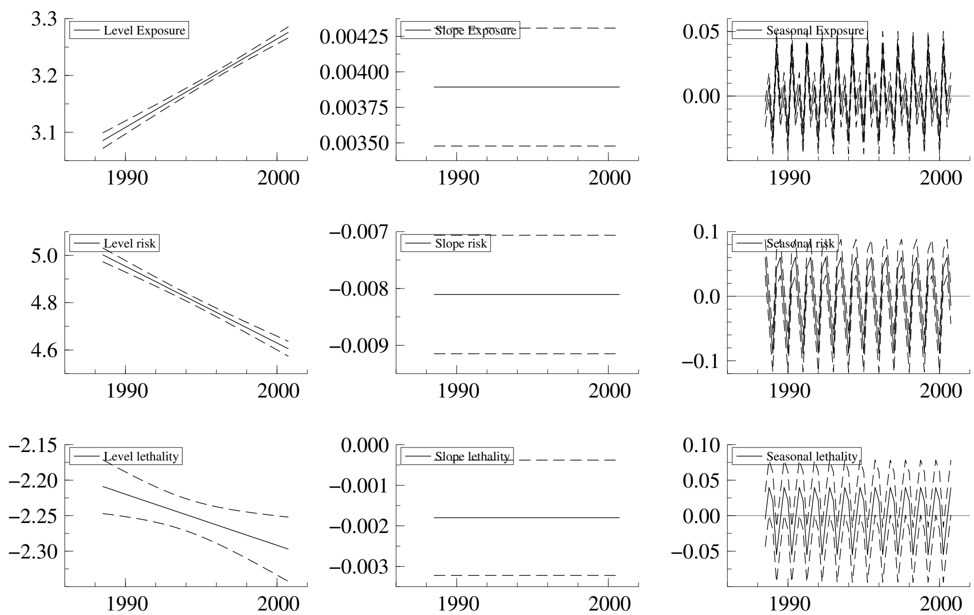

Figure 3.5. Smoothed trend, slope and seasonal components of multivariate linear regression for the whole of the Netherlands, including 95% confidence limits.

As Figure 3.5 clearly indicates, the trends for exposure, accident risk, and fatality risk are all now simple regression lines with a fixed slope and intercept, and the seasonal components no longer change over time. In contrast with the previous analysis, the assumption of independence of the standardised prediction errors is now violated for the KSI accidents series, and the assumption of homoscedasticity is now violated for the vehicle kilometres series and for the KSI accidents series. As a result, in Figure 3.4 more observations fall outside the 95% confidence interval for the smoothed KSI accident predictions than in Figure 3.1.

At convergence the value of the log-likelihood function for the completely deterministic model equals 201.429. Applying the likelihood ratio test to the results obtained with the nested stochastic and deterministic models, we find that

$$2(\log(L_u) - \log(L_r)) = 2(219.746 - 201.429) = 36.634\,.$$

Since this value of the likelihood ratio test is larger than

$$\mathrm{X}^2_{(0.05;18)} = 28.869\,,$$

where (24 - 6) = 18 is the number of restrictions, the improvement in fit of the stochastic model compared to the deterministic model is significant with $p < 0.05$.

The forecasts obtained with the two models are shown in Figures 3.6 and 3.7.

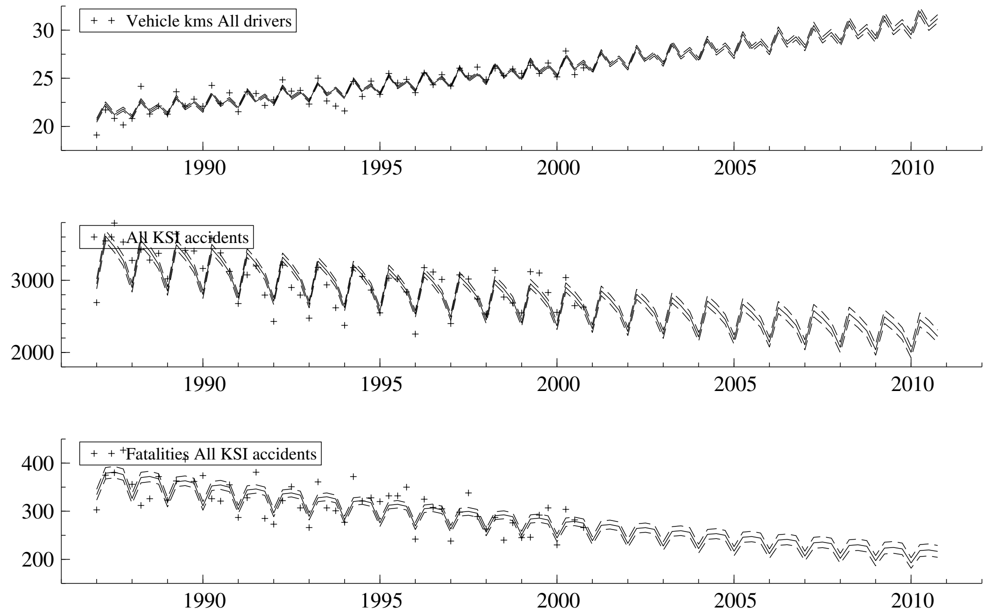

Figure 3.6. Forecasts of deterministic three-level model applied to the whole of the Netherlands, including 68% confidence limits.

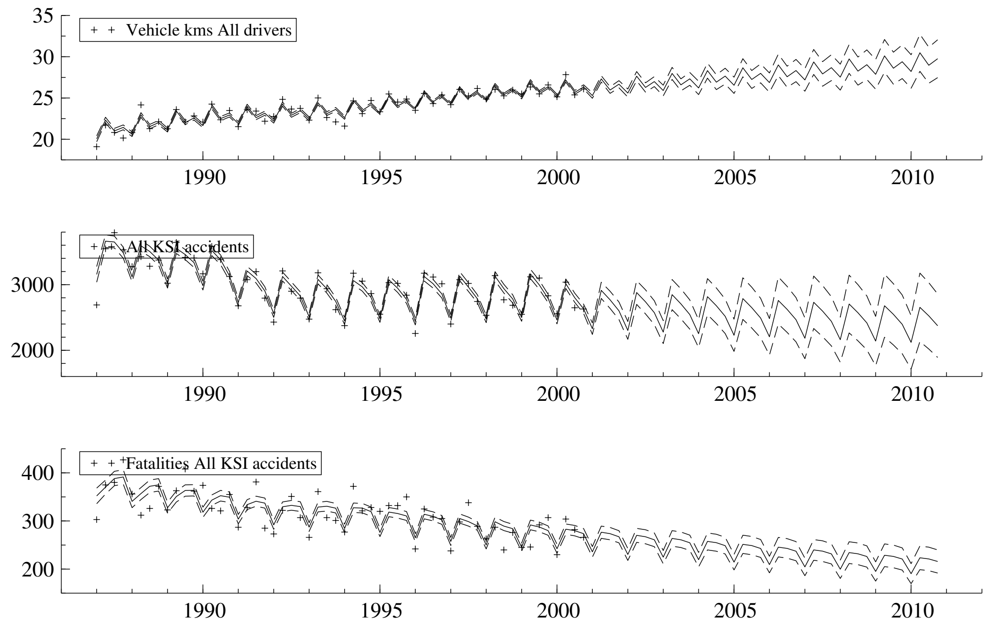

Figure 3.7. Forecasts of stochastic three-level model applied to the whole of the Netherlands, including 68% confidence limits.

Just as in the analyses of the French data, the confidence limits surrounding the forecasts in Figure 3.7 are also much more conservative than in Figure 3.6, indicating that forecasts obtained with the multivariate stochastic model result in a much more realistic level of uncertainty concerning future developments in road safety.

Chapter 4 - French road safety developments disaggregated by road type

Here we present the results of the same three-level analysis as performed in Section 2 for the whole of France. In this section, however, the analyses are applied to two networks: main roads and motorways, for which the number of vehicle kilometres driven are available.

For each of the two networks, we simultaneously analyse quarterly data for the period 1987-2000 concerning:
- the number of vehicle kilometres driven (from which the unobserved exposure is derived),
- the total number of injury accidents (from which the unobserved accident risk is derived),
- the total number of people being killed (from which the unobserved fatality risk is derived).

	Main roads	Motorways	**Main network**	**Whole France**
Injury accidents	11 807	5 910	**17 717**	**124 387**
monthly average	984	493	**1 476**	**10 366**
% whole France	*9,49%*	*4,75%*	***14,24%***	
Fatalities	1 928	471	**2 399**	**8 437**
monthly average	161	39	**200**	**703**
% whole France	*22,85%*	*5,58%*	***28,43%***	
Traffic volume (10^8 veh-km)	886	969	**1 855**	
monthly average	74	81	**155**	
% whole France			***about 30%***	
Network length (km)	24 000	8 763	32 763	

Table 4.1. The French data for 1998

As indicated in Table 4.1 and in Appendix 4, the French main network, on which the number of vehicle kilometres driven is measured by sensors, handles about one third of the traffic volume of the whole of France, and around 14% of injury accidents and 28% of fatalities (data for the year 1998). However, although the traffic volume is equally shared between main roads and motorways, this is not the case for the number of injury accidents and the number of fatalities, and as a consequence for the accident risk and fatality risk on each network. Moreover, as it will be seen in the next subsections, the accident risk and fatality risk developments for the period 1997-2000 show differences between the two networks, and with the whole of France.

4.1. Main roads

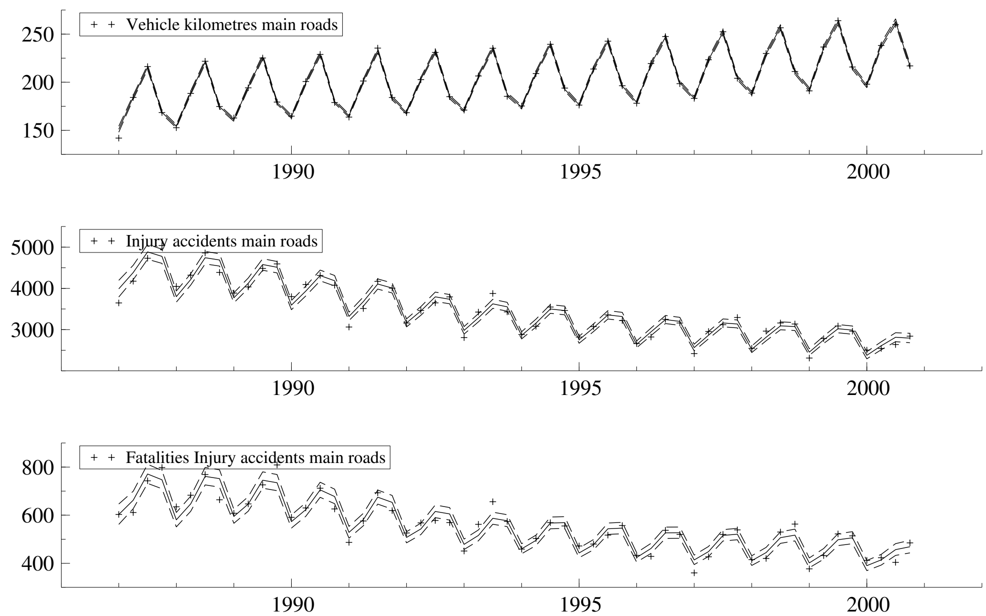

Figure 4.1. Smoothed predictions for quarterly vehicle kilometres, KSI accidents, and fatalities on French main roads, including 95% confidence limits.

In contrast with the proxy oil sales for exposure used in the analysis for the whole of France, the number of vehicle kilometres on French main roads is described quite accurately, as reflected in the fact that the 95% confidence limits for the smoothed model predictions for this variable are now much smaller than the confidence limits for oil sales in Figure 2.1.

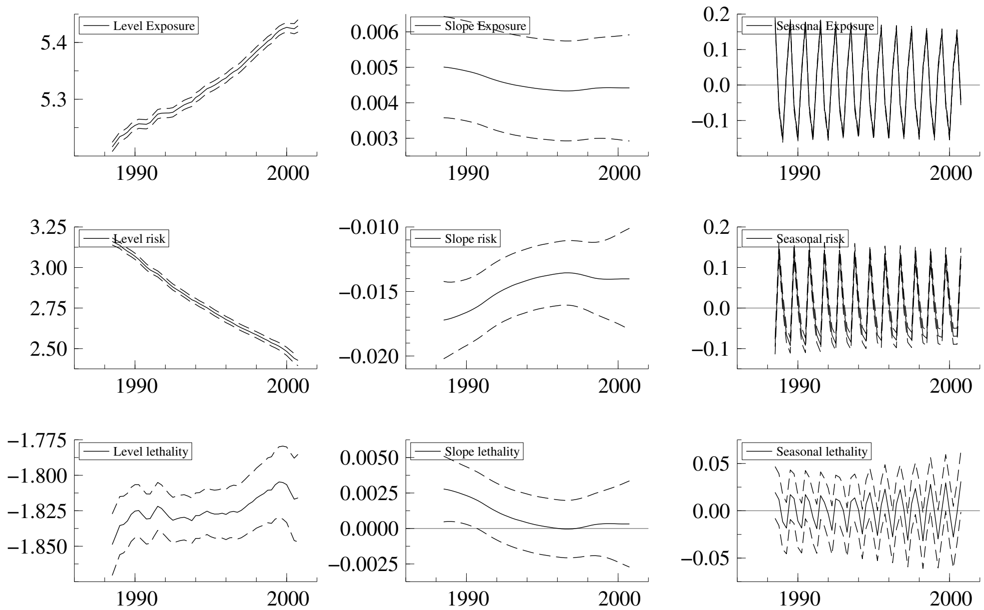

Figure 4.2. Smoothed trend, slope, and seasonal components for French main roads, including 95% confidence limits.

The general trend developments of exposure and of risk are, on the French main roads, about the same as for the whole of France. Nevertheless, the development in fatality risk is slightly different from the whole of France in that it is almost continuously increasing.

4.2. Motorways

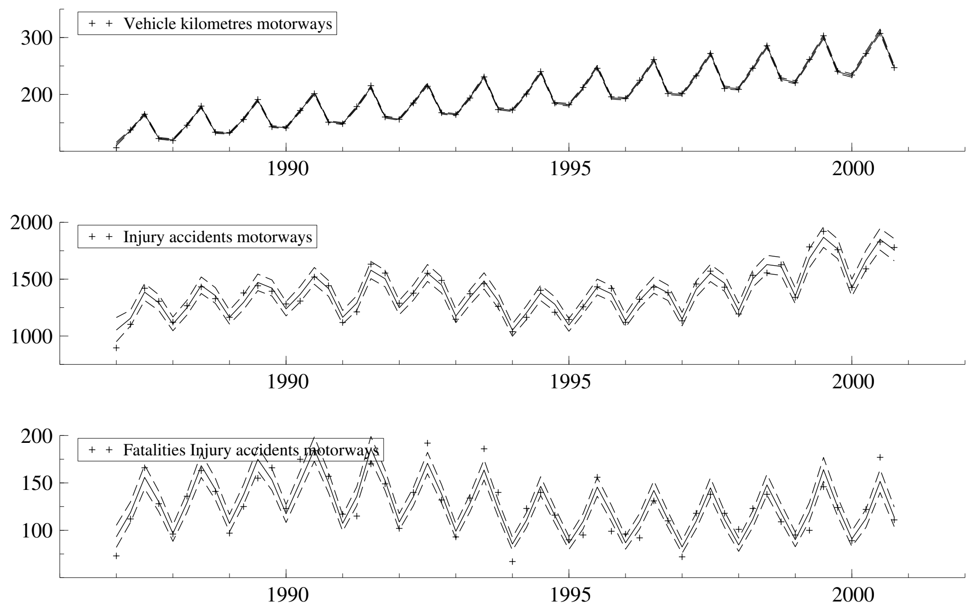

Figure 4.3. Smoothed predictions for quarterly vehicle kilometres, KSI accidents, and fatalities on French motorways, including 95% confidence limits.

Again we see that the number of vehicle kilometres on French motorways is described quite accurately since the 95% confidence limits of the smoothed model predictions for this variable are now also much smaller than for the oil sales used as a proxy for exposure in the analysis for the whole of France.

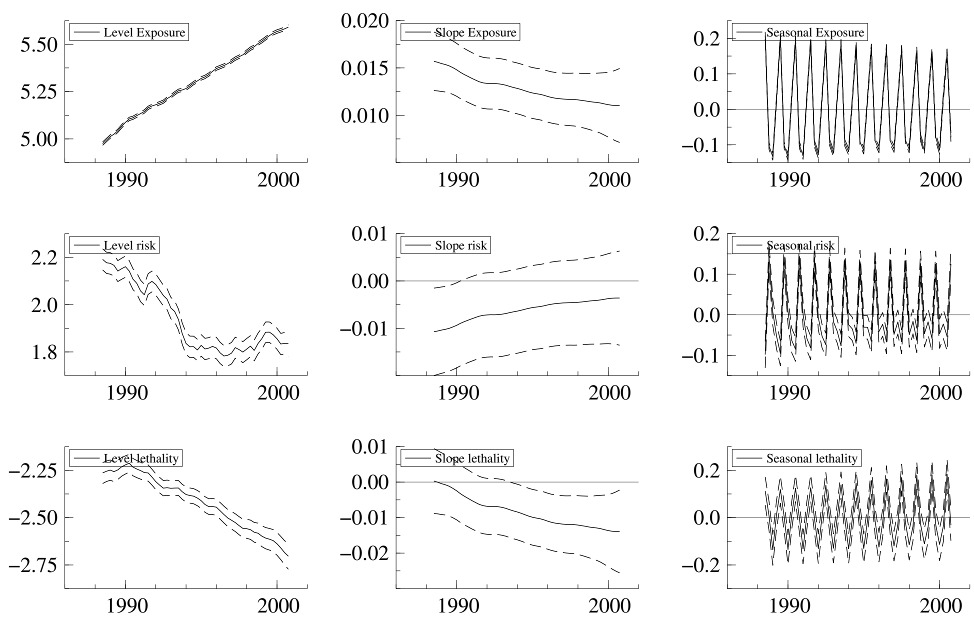

Figure 4.4. Smoothed trend, slope, and seasonal components for French motorways, including 95% confidence limits.

On the French motorways, the injury accident risk decreased continuously just as for the whole of France. In contrast, the fatality risk decreased continuously on the French motorways, which was not the case for the whole of France.

Chapter 5 - Dutch road safety developments disaggregated by type of KSI accident

Here we present some results of applying a multivariate model to different types of Dutch accident data with people killed and seriously injured (KSI). We have performed five separate analyses of quarterly data for the years 1994-2006 on all KSI accidents and on car-car KSI accidents in the Netherlands. In both cases, five types of accidents were considered: all accidents (see also Section 3), side accidents, frontal accidents, rear-end accidents, and single accidents. Single accidents also include collisions with obstacles like trees, etc.

As before, for each of these accident types the following three levels of road safety were simultaneously analysed:
- vehicle kilometres (from which the unobserved exposure is derived),
- the number of KSI accidents (from which the unobserved severe accident risk is derived),
- the number of fatalities (from which the unobserved fatality risk is derived).

5.1. All KSI accidents

All estimated trends in the analyses are presented in one figure in order to simplify the comparison between the different types of KSI accidents. First, the trends for the unobserved exposure are presented in Figure 5.1.

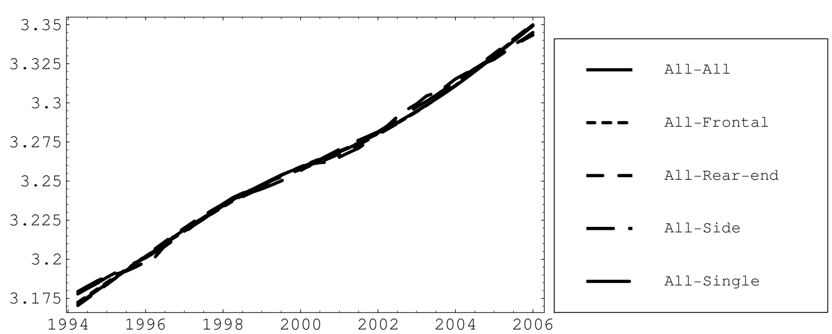

Figure 5.1. Trends of the latent variable for total exposure.

As Figure 5.1. shows these trends are very similar for the different types of KSI accidents, and are all continuously increasing. The trends differ slightly due to the fact that uncertainty in the traffic volume data is considered, and different dependent variables due to fact that different accident types are considered.

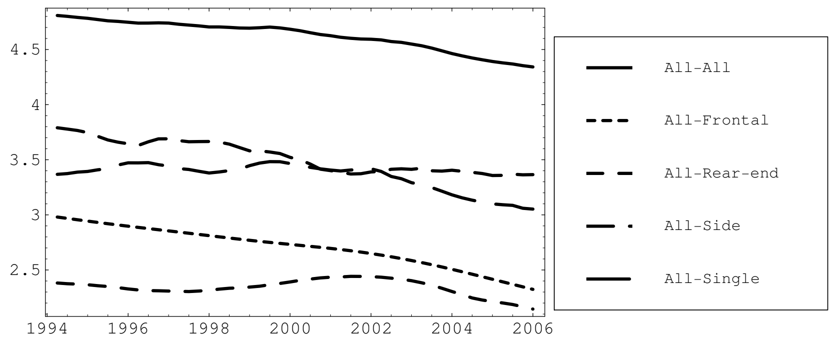

Figure 5.2. Trends of latent variable for total accident risk: KSI accidents per driver kilometre.

On the whole, the risks of becoming involved in the different types of KSI accident were continuously decreasing, except for rear-end KSI accidents whose risk was slightly increasing in the years 1999 and 2000. Also, throughout the period 1994-2006 the risk of being involved in a KSI accident was largest for side and single accidents, smaller for frontal, and smallest for rear-end accidents. Note that the risk of all accidents in Figure 5.2 is the sum of the risks of the individual types of accidents, and that the figure contains the log of these risks.

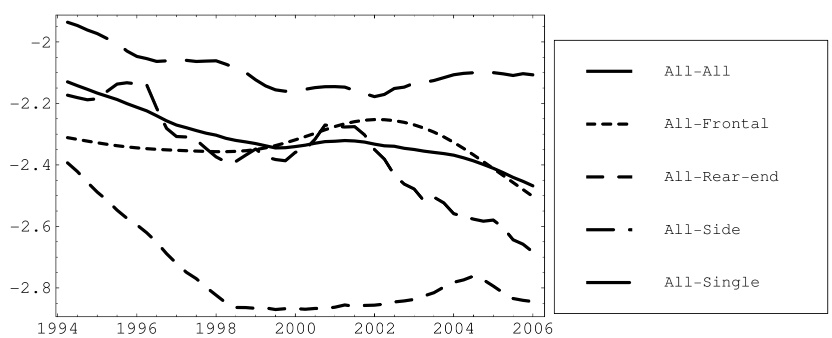

Figure 5.3. Trends of latent variable for total severity risk: fatalities per KSI accident.

The fatality risk was the smallest for rear-end accidents (see Figure 5.3). This can be explained by the fact that the speed of impact is generally smallest for this type of collision. Apart from speed, the crushable zone is also the largest in this situation, at least for cars and trucks. The fast decrease in this type of risk between 1994 and 1998 may be the result of increasing congestion on Dutch roads.

The fatality risk was higher for frontal and side accidents. This can be explained by the fact that the speed of impact is generally larger in frontal and side accidents than in rear-end accidents. Moreover, the amount of protection against a collision from the side in vehicles is particularly low.

As Figure 5.3 indicates the risk of being killed is the largest for single accidents (even though these accidents all only involve one vehicle). This could be explained by groups of young people circulating in traffic during weekend nights.

To simplify the graphs of these disaggregated developments the confidence intervals associated with all these trends have not been shown. However, the estimation error variances associated with all these trends (and from which confidence intervals can be constructed under the assumption of normality) can be used to test whether the trends for the accident types are significantly different from each other or not.

5.2. Car versus car KSI accidents

The same analyses as performed above were also applied to quarterly car versus car KSI accident data in the Netherlands. In these analyses car-all implies car-car accidents only (i.e., without single car accidents). However, for purposes of comparison we have also added the results of the three-level analysis of all single KSI car accidents. The related results, in terms of trends, are summarised in the following figures.

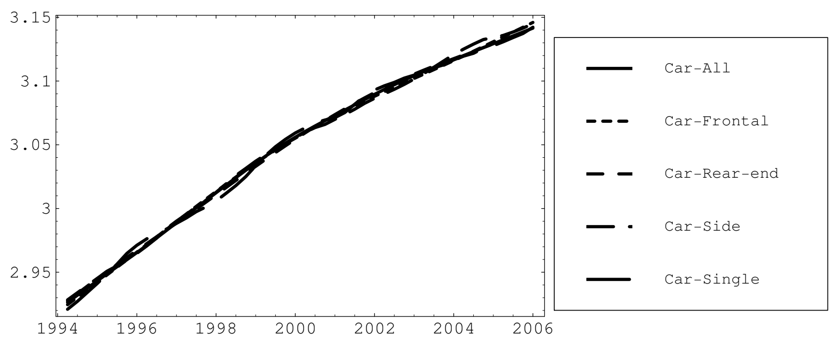

Figure 5.4. Trends of latent variable for Car-car exposure.

Again the trends for the *latent variable for* exposure for all these types of car accidents are very similar.

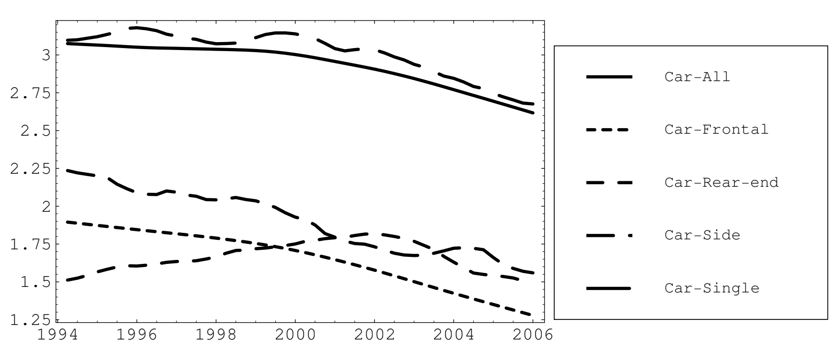

Figure 5.5. Trends of latent variable for Car-car accident risk: KSI accidents per driver kilometre.

Figure 5.5 shows that the risk of being involved in a single car KSI accident was approximately equal to the risk of being involved in a car-car KSI accident (irrespective of type). Moreover the latter two risks were both higher than those for any type of car-car accidents.

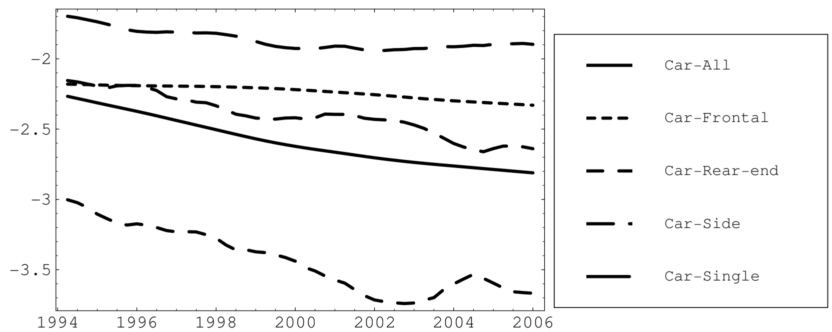

Figure 5.6. Trends of latent variable for Car-car fatality risk: fatalities per KSI accident.

Inspection of Figure 5.6 shows that the fatality risk was largest for single car accidents, while it was smallest for rear-end car-car accidents. Though yielding a less severe outcome than single car accidents, frontal and side car-car accidents are still clearly more lethal than rear-end car-car accidents, for reasons already given in the previous section.

Chapter 6 - Conclusion

In this deliverable, a multivariate time series model was proposed for the analysis of SafetyNet data using state space methods. The main objective of these time series analysis techniques was to provide estimations of the relevant following variables, namely: exposure to risk, accident risk (the risk an injury accident happens), and fatality risk (the risk of being killed in an injury accident), estimated through their unobserved components: their trend (level and slope) and their seasonals.

We applied the multivariate model on national quarterly time series data from France and from the Netherlands. Results of these two analyses were presented and compared. It was shown that the stochastic multivariate time series model not only yielded a significant improvement in goodness-of-fit compared to the standard vectorial linear regression model, but also that all model assumptions were satisfied in the former model while several assumptions were violated in the standard approach.

We also applied the multivariate model to two types of French roads, and to several types of Dutch KSI accidents. Results of these analyses were also presented.

Differences between exposure, accident risk and fatality risk developments were demonstrated, when applied to aggregated and disaggregated road safety data.

This multivariate model can be used on aggregate data of European countries with two main objectives: for monitoring purposes, on the one hand and for comparison purposes, on the other hand.

To describe the recent developments in the trends of exposure, accident risk and fatality risk requires to correct these variables for their seasonal component: it is for instance worth noting that the seasonal pattern of this seasonal component changes with time, which appears particularly clearly in the case of the Dutch aggregate data (Seasonal Risk in Figure 3.2).

To compare developments in the trends of these variables, between types of networks, types of accidents, or between European countries, will allow to explain, as a second step, the differences in the developments of the levels and slopes of those stochastic, assumed to be locally linear, trends. These differences in the development of the slopes and levels of the fatality risk, for instance, appear particularly clearly in the case of the French data disaggregated on main roads and motorways (Compare Figure 4.2 to Figure 4.4).

It should be noted that, the accident risk may is not comparable between France and the Netherlands, due to the differences in exposure to risk measures used (vehicle kilometres versus oil sales). It is for this reason that available data resulting from SafetyNet WP2 will become very important in the future. Furthermore, differences in the definition of fatalities render comparisons difficult as well. It is assumed that the outputsresults of the SafetyNet WP1 project will reduce such problems. Both issues make a genuine comparison between road safety developments in France and the Netherlands difficult.

An obvious extension of the descriptive multivariate models presented in this deliverable would be to investigate and evaluate the possible effects of explanatory and intervention variables on the modelled developments in exposure, accident risk and fatality risk. This for example could be done with time series data on the safety performance indicators collected in WP3 of the SafetyNet project.

Another interesting topic for further research would be to investigate whether the estimated variance-covariance structures of the three-level model contain common components (either in levels, or in slopes, or in seasonals), in the end allowing to apply the model as a form of dynamic principal components analysis including all the usual interpretations possible in a static principal components analysis. One application of this may be to identify common trends in road safety among member states of the European Union.

References

Bergel, R. (1997). *Multivariate modelling of road risk on the French main road network* Proceedings of the 25th European Transport Annual Meeting PTRC, Brunel University, Uxbridge, London, 1-5 september 1997. Also partly in COST 329 (2004).

Bijleveld, F.D., Commandeur, J.J.F., Gould, P.G., and Koopman, S.J. (2005). Model-based measurement of latent risk in time series with applications. *TI 2005-118/4. Amsterdam: Tinbergen Institute (www.tinbergen.nl).*

Bijleveld, F.D., Commandeur, J.J.F., Gould, P.G., and Koopman, S.J. (forthcoming). Model-based measurement of latent risk in time series with applications. *Journal of the Royal Statistical Society Series A*.

COST 329 (2004). *Models for traffic and safety development and interventions. European Commission*. Directorate general for Transport, Brussels.

Doornik, J. A. (2001). *Object-oriented matrix programming using Ox 3.0*. London: Timberlake Consultants Press.

Durbin, J. & Koopman, S.J. (2001). *Time series analysis by state space methods.* Oxford: Oxford University Press.

Gaudry M. (1984). *DRAG, un modèle de la Demande Routière, des Accidents et de leur Gravité, appliqué au Québec de 1956 à 1982*, Publication 359, Centre de Recherche sur les Transports, Université de Montréal.

Gaudry M. and Lassarre S. (Ed.) (2000). *Structural Road Accident Models - The International DRAG Family*, Pergamon.

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge: Cambridge University Press.

Koopman, S. J., Shephard, N., & Doornik, J. A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal*, Vol. 2, p.113-166.

Lassarre, S. (2001). Analysis of progress in road safety in ten European countries. *Accident Analysis and Prevention*, Vol. 33, 743-741.

Yannis G and al. (2005). *State of the Art Report on Risk and Exposure Data*. Safetynet Deliverable 2.1.

Appendix 1

This Appendix discusses the general non-seasonal case.
By adding stochastic seasonal components to this multivariate local linear trend framework for annual data it can easily be extended to the modelling of quarterly or monthly data.
Define

$$y_t = \begin{pmatrix} y_t^{(1)} \\ y_t^{(2)} \\ y_t^{(3)} \end{pmatrix} = \begin{pmatrix} \log V_t \\ \log A_t \\ \log F_t \end{pmatrix},$$

where V_t is the observed number of vehicle kilometres, A_t is the observed number of injury accidents, and F_t is the observed number of fatalities at time points $t = 1, ..., n$.
Also define

$$\alpha_t = \begin{pmatrix} \mu_t^{(1)} \\ \mu_t^{(2)} \\ \mu_t^{(3)} \\ \nu_t^{(1)} \\ \nu_t^{(2)} \\ \nu_t^{(3)} \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \\ \xi_t^{(3)} \\ \zeta_t^{(1)} \\ \zeta_t^{(2)} \\ \zeta_t^{(3)} \end{pmatrix}, \quad T_t = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad R_t = I,$$

$$Z_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad H_t = \begin{bmatrix} \sigma^2_{\varepsilon^{(1)}} & \mathrm{cov}(\varepsilon^{(1)},\varepsilon^{(2)}) & \mathrm{cov}(\varepsilon^{(1)},\varepsilon^{(3)}) \\ \mathrm{cov}(\varepsilon^{(1)},\varepsilon^{(2)}) & \sigma^2_{\varepsilon^{(2)}} & \mathrm{cov}(\varepsilon^{(2)},\varepsilon^{(3)}) \\ \mathrm{cov}(\varepsilon^{(1)},\varepsilon^{(3)}) & \mathrm{cov}(\varepsilon^{(2)},\varepsilon^{(3)}) & \sigma^2_{\varepsilon^{(3)}} \end{bmatrix},$$

and

$$Q_t = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$$

with

$$B = \begin{bmatrix} \sigma^2_{\xi^{(1)}} & \mathrm{cov}(\xi^{(1)},\xi^{(2)}) & \mathrm{cov}(\xi^{(1)},\xi^{(3)}) \\ \mathrm{cov}(\xi^{(1)},\xi^{(2)}) & \sigma^2_{\xi^{(2)}} & \mathrm{cov}(\xi^{(2)},\xi^{(3)}) \\ \mathrm{cov}(\xi^{(1)},\xi^{(3)}) & \mathrm{cov}(\xi^{(2)},\xi^{(3)}) & \sigma^2_{\xi^{(3)}} \end{bmatrix}$$

and

$$C = \begin{bmatrix} \sigma^2_{\zeta^{(1)}} & \text{cov}(\zeta^{(1)}, \zeta^{(2)}) & \text{cov}(\zeta^{(1)}, \zeta^{(3)}) \\ \text{cov}(\zeta^{(1)}, \zeta^{(2)}) & \sigma^2_{\zeta^{(2)}} & \text{cov}(\zeta^{(2)}, \zeta^{(3)}) \\ \text{cov}(\zeta^{(1)}, \zeta^{(3)}) & \text{cov}(\zeta^{(2)}, \zeta^{(3)}) & \sigma^2_{\zeta^{(3)}} \end{bmatrix},$$

where $\{\mu_t^{(1)}, \mu_t^{(2)}, \mu_t^{(3)}\}$ denote the level components and $\{\nu_t^{(1)}, \nu_t^{(2)}, \nu_t^{(3)}\}$ denote the slope components of the trends for exposure, accident risk and fatality risk respectively.

The three-level model discussed in this deliverable is a special case of the general state space methods for the analysis of time series (Harvey, 1989; Durbin and Koopman, 2001). In matrix algebra, all state methods can generally be written as

$$y_t = Z_t \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, H_t) \tag{1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \qquad \eta_t \sim NID(0, Q_t) \tag{2}$$

for t = 1, ..., n, where (1) is called the observation or measurement equation (and ε_t is an irregular component consisting of observation errors or disturbances), and (2) is called the state equation (and η_t contains the state disturbances). The present model for evaluating different levels of road safety is a multivariate local linear trend model.

With the above definitions it can be verified that we obtain the following model. First writing out (1) in scalar notation yields the following three observation equations:

$$\begin{aligned} y_t^{(1)} &= \mu_t^{(1)} + \varepsilon_t^{(1)} \\ y_t^{(2)} &= \mu_t^{(1)} + \mu_t^{(2)} + \varepsilon_t^{(2)} \\ y_t^{(3)} &= \mu_t^{(1)} + \mu_t^{(2)} + \mu_t^{(3)} + \varepsilon_t^{(3)} \end{aligned} \quad , \tag{3}$$

while working out (2) in scalar notation results in the following six state equations:

$$
\begin{aligned}
\mu_{t+1}^{(1)} &= \mu_t^{(1)} + \nu_t^{(1)} + \xi_t^{(1)} \\
\mu_{t+1}^{(2)} &= \mu_t^{(2)} + \nu_t^{(2)} + \xi_t^{(2)} \\
\mu_{t+1}^{(3)} &= \mu_t^{(3)} + \nu_t^{(3)} + \xi_t^{(3)} \\
\nu_{t+1}^{(1)} &= \nu_t^{(1)} + \zeta_t^{(1)} \\
\nu_{t+1}^{(2)} &= \nu_t^{(2)} + \zeta_t^{(2)} \\
\nu_{t+1}^{(3)} &= \nu_t^{(3)} + \zeta_t^{(3)}
\end{aligned}
\quad . \tag{4}
$$

By adding stochastic seasonal components to this multivariate local linear trend framework for annual data it can easily be extended to the modelling of quarterly or monthly data.

Appendix 2

The estimated variance matrix of the observation disturbances for the whole of France is:

	oil sales	injury accidents	fatalities
oil sales	9.05735E-5	0.000155929	0.000135539
injury accidents	0.000155929	0.000596208	0.000445838
fatalities	0.000135539	0.000445838	0.000911089

The estimated variance matrices of the state disturbances for the whole of France are equal to:

	exposure	accident risk	fatality risk
level components			
exposure	0.0000368091	-0.0000103146	0.0000361062
accident risk	-0.0000103146	0.000318722	0.000141376
fatality risk	0.0000361062	0.000141376	0.000108083
slope components			
exposure	7.01823E-7	-0.00000112504	0.00000106196
accident risk	-0.00000112504	0.00000180348	-0.00000170236
fatality risk	0.00000106196	-0.00000170236	0.00000106196
seasonal components			
exposure	3.98743E-7	-2.2089E-7	0.00000148649
accident risk	-2.2089E-7	1.22366E-7	-8.23466E-7
fatality risk	0.00000148649	-8.23466E-7	0.00000554155

Appendix 3

The estimated variance matrix of the observation disturbances for the whole of the Netherlands is:

	motor vehicle kms	injury accidents	fatalities
motor vehicle kms	0.000241432	0.000184007	0.000540755
injury accidents	0.000184007	0.00161536	0.0020346
fatalities	0.000540755	0.0020346	0.0082125

The estimated variance matrices of the state disturbances for the whole of the Netherlands are equal to:

	exposure	accident risk	fatality risk
level components			
exposure	2.67605E-5	4.02208E-5	-8.8639E-5
accident risk	4.02208E-5	0.000285414	-0.000230567
fatality risk	-8.8639E-5	-0.000230567	0.000335721
slope components			
exposure	9.41423E-8	-2.70138E-7	1.46211E-7
accident risk	-2.70138E-7	7.75149E-7	-4.19546E-7
fatality risk	1.46211E-7	-4.19546E-7	2.27078E-7
seasonal components			
exposure	6.46107E-6	-2.26173E-5	-8.50118E-6
accident risk	-2.26173E-5	7.91729E-5	2.97588E-5
fatality risk	-8.50118E-6	2.97588E-5	1.11855E-5

Appendix 4

The French data for the year 1998 are presented in the following table. The French main network on which the number of vehicle driven are measured by sensors, handles about one third of the traffic volume of the whole of France, and around 14% of injury accidents and 28% of fatalities (data for the year 1998). It is constituted of main roads on the one hand, and of (free and public) motorways on the other hand, on which the traffic volume is about equally shared - which is not the case for the number of injury accidents and fatalities, and as a consequence for the accident risk and fatality risk.

	Main roads	Motorways	**Main network**	Secondary roads	Urban roads	**Whole France**
Injury accidents	11 807	5 910	**17 717**	36 278	70 392	**124 387**
monthly average	984	493	**1 476**	3023	5866	**10 366**
% whole France	*9,49%*	*4,75%*	***14,24%***	*29,16%*	*56,59%*	
Fatalities	1 928	471	**2 399**	4 373	1 665	**8 437**
monthly average	161	39	**200**	364	139	**703**
% whole France	*22,85%*	*5,58%*	***28,43%***	*51,83%*	*19,73%*	
Traffic volume (10^8 veh-km)	886	969	**1 855**			
monthly average	74	81	**155**			
% whole France			***about 30%***			
Network length (km)	24 000	8 763	32 763			