8. Conclusions

This thesis presents a set of comprehensive studies into time series analysis for aggregated road safety data, such as accident counts and victim counts. In particular, the number of fatalities and serious injuries is closely monitored by government agencies and the public, and its relevance to society is not disputed. Much research is conducted into how road safety can be improved. To that end it is often attempted to explain changes in road safety statistics by factors (or changes in) such as exposure, policy, driving under the influence of alcohol, speeding by drivers and infrastructural measures. Some factors such as regulations, traffic law and policy can be directly observed (although compliance with regulations, traffic law and policy may not). Other factors can be observed in theory but in practice their measurement is either difficult or very expensive. Examples of such factors are exposure, which can be measured using surveys and vehicle counting systems, and driving under the influence of alcohol, which can be measured using road side surveys. Finally, some factors are even harder to observe such as driver skill or experience. Data obtained from diverse sources as described above are likely to differ in accuracy, which may complicate statistical analysis.

Another complicating factor in road safety time series analysis is that no unique measure of road safety is available. Usually road safety is measured in terms of the number of accidents or the number of victims. Although in practice the situation is more complicated, some road safety measures may affect either accident occurrence or accident severity. When a study is performed measuring the effect of a policy intervention on road safety that should mainly affect accident occurrence, the development of the number of accidents of a relevant kind would be studied. On the other hand, if the policy intervention should mainly affect accident severity, the development of the number of victims would be studied. If possible, an analysis is performed on both a type of victim for which the count should be affected by the policy intervention, and a type of victim for which the count is not affected by the policy intervention. The development of both victim types should otherwise be as similar as possible. If a reduction in the number of victims is indeed identified it is important to confirm that the number of victims is reduced because accident severity is reduced, not because the number of accidents is reduced. Ideally, changes should not be found for the type of victim of which the count should not be affected by the policy intervention. Unfortunately, it is not likely that both conditions are met when the number of victims is studied for a longer period of time. The possibility has to be considered that other influences have affected road safety. These influ-
ences themselves may need to be modelled and need not be fully independent of the policy intervention originally considered. To accommodate this case it is important to jointly model influences on joint dependent variables, notably accident counts and victim counts.

In this thesis a novel approach to road safety time series analysis of developments of aggregated road safety data is presented intended to improve options and yield more reliable statistical analysis compared to commonly used alternatives. The approach is based on multivariate heteroscedastic structural time series models and addresses many of the issues in road safety time series analysis, and all of the above described issues. Currently applied models may only partly treat these issues.

The combination of three basic aspects of the approach presented in this thesis allow for the improvements to road safety time series models. These three aspects are:

- **The use of structural components.** The time series models are constructed using interpretable structural components, such as exposure, risk and severity components. Also registration rate, seat belt use percentage, percentage of drivers exceeding the legal blood alcohol concentration limit, speeding behaviour and other important aspects of road safety may be represented using structural components. The structural components may follow trends and seasonal patterns. The benefits of using interpretable components become apparent when a single component can be related to more than one dependent variable. Furthermore, these models allow the researcher to distinguish external effects into effects that affect road safety or its important components (for instance interventions) or effects that affect how road safety is observed (for instance changes in registration rate). Furthermore explanatory variables can be modelled to have an effect on specific components of road safety.

- **Multivariate dependent variables:** both accident and victim counts can be included in one model. Explanatory variables can be included in a traditional way. In addition, explanatory variables measured with observation error can be included as well, where a structural component can be considered an estimate of the true value (without observation error). For instance if the explanatory variable is seat belt use, data may be obtained from relatively small road side surveys, not necessarily available for all observations. If seat belt use can be considered relatively constant, the structural component can be the average of all observations. This
structural component could then be used for all observations, including observations where no survey result is available. If seat belt use cannot be considered constant, a trend can be considered, still allowing the model to use all observations.

- Heteroscedastic structure of errors: the models can treat observations which accuracy may vary over time. The observations may also be unavailable at a few time points. The models can treat cases where variables differ in accuracy. For instance, traffic volume data for cars may be more accurate than traffic volume data for motorcycles (the relative error for the total traffic volume of cars is about 2–3% in the survey of 2003, the relative error for the total traffic volume of motorcycles is almost 30% in the survey of 2003). Yet both variables may be included in the same model.

It is the combination of these properties in one model and the use of shared structural components that makes the model particularly attractive to road safety time series analysis. Considering that a longer period of time is studied, where conditions change over time, the following practical benefits can be mentioned:

- The multivariate nature and the heteroscedastic nature of the models are utilised to account for covariance among dependent variables. It is known that the number of victims is dependent on the number of accidents. Therefore models including both accident counts and victim counts should account for their covariance. Similarly, the error in travel data from surveys can be correlated. The covariance between accident counts and victim counts is successfully used in both applications of Chapter 3. The error in travel data from surveys is used in the applications of Chapter 3, the Dutch application in Chapter 5 and the road safety application in Chapter 7.

- Structural components allow for a limited level of verification. The development of structural components can be compared with secondary information. However, if sufficient secondary information is available, it is probably better to include that information in the model. A successful application of this approach is in Chapter 6 where the development of exposure outside urban areas is compared to an estimate of traffic volume constructed from road length data and traffic intensity data.

- Structural components can be shared among many dependent variables. In the road safety application in Chapter 7, weather conditions measured
by ten weather stations is related to one single structural component. If the weather stations agree, the value of the structural component is accurately known. On the other hand, if the weather stations disagree, that is, the weather pattern differs among the weather stations, the value of the structural component is not accurately known. The possibility of sharing components is also utilised in Chapter 3 where a structural component representing the number of victims per accident is shared among police recorded victims and an estimate of the so called true number of victims. This component is used to improve the estimate of the true number of accidents, which cannot be inferred from hospital records.

- The approach allows for other observation error distributions than Gaussian. Even combinations of other observation error distributions are provided for by the modelling approach presented in this thesis. Notably, an example is given using both multiple Gaussian distributed observation errors and multiple Poisson distributed dependent variables. This feature is used in the road safety application in Chapter 7, where Poisson distributed accident counts are analysed for both dry and wet weather conditions as well as Gaussian distributed traffic volume data and weather station data.

- The approach allows correlation among so called innovations of structural components. When innovations of structural components are allowed to correlate (and are correlated), the developments of structural components mutually affect each other. This is relevant when components represent phenomena that may affect each other, such as exposure and risk. Also the development of accident severity may affect the occurrence of accidents exceeding a certain severity level. The fact that exposure can affect risk is used in almost all examples in this thesis (eliminating the need for a coefficient for exposure), while the fact that accident severity can affect occurrence of severe accidents is used in the first application of Chapter 3.

In the model applications presented in this thesis the definition of exposure used is always proportional to traffic volume. This restriction is not fundamental however. The approach presented in this thesis can also be applied to models assuming a nonlinear relation, and using other exposure measures than traffic volume. Furthermore, the dynamic relations assumed in this these are derived from local linear trend models. Although smooth developments may be approximated satisfactorily in practice using local linear trend models, this need not always be the case. In particular the local level approximation to
the development of the fraction of traffic with precipitation in Chapter 7 may be improved upon. As this thesis is aimed at providing better and statistically more reliable options for time series analysis of road safety data rather than generating empirical results, the improvement of such models is considered outside the scope of this thesis.

The models considered in this thesis all assume a linear dynamic relation, and Gaussian state disturbances. These assumptions may not hold in all cases, although the approximately linear local linear trend models appear to function quite well in practice. Sufficient empirical evidence is given in this thesis that shows the effectiveness of the new proposed methodology of time series analysis for traffic safety data. It is planned to develop the methodology further into higher dimensions and into more realistic models for traffic safety. In this thesis the key contributions of the new approach are reported.